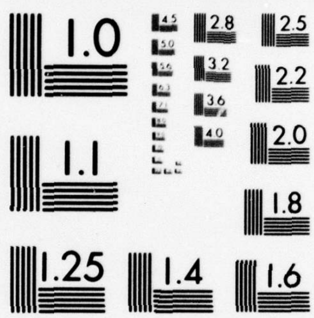
AD-A065 289 MICHIGAN UNIV ANN ARBOR INST FOR SOCIAL RESEARCH F/G 5/11
ORGANIZATIONAL DIAGNOSIS: ISSUES AND METHODS.(U)
FEB 79 A S DAVENPORT, D G BOWERS N00014-77-C-0096

UNCLASSIFIED NL

1 OF 2
AD
A065289



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD A0 65289

DDC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

12

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Organizational Diagnosis: Issues and Methods		5. TYPE OF REPORT & PERIOD COVERED (9) Final Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Alan S. Davenport, David G. Bowers		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-77-C-0096
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Social Research, University of Michigan, Ann Arbor, Michigan		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Manpower Research & Development Program, Office of Naval Research, Arlington, VA		12. REPORT DATE (11) February, 1979
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

180 300

LEVEL

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

(12) 147 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

D D C
RECEIVED
MAR 6 1979
C

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Bayes	Decision Tree	Likert
CANOPUS	Diagnosis	Military patterns
Change Agent	Discriminant Function	Misclassification
Classification	Distance Function	Model
Classification Criteria	Intervention	Organizational Diagnosis OVER

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents different approaches to data based organizational diagnosis (in particular, distance function, discriminant function, decision tree, and a Bayesian approach) and their effectiveness when applied to approximately 6,000 work groups. In particular, several different criteria are used to evaluate the classification techniques: proportion of correct classification, ability to reproduce the typology, a zero-one closeness measure, severity of misclassification, the information required to make classifications, the amount of data required to develop the procedure, and the ease of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

180 300 79 03 02 031 SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 19 continued - Key Words

Organizational Functioning
Organizational Theory
Profile Scoring
Standardized Instrument
Systematic Diagnosis
Survey of Organizations
Treatment
Typology

Block 20 - Abstract continued

implementing the procedure. Discussion of these issues, and, as a consequence, suggestions for further investigation are included.

FINAL TECHNICAL REPORT

February, 1979

ORGANIZATIONAL DIAGNOSIS:

ISSUES AND METHODS

By

Alan S. Davenport

David G. Bowers

Institute for Social Research
University of Michigan
Ann Arbor, Michigan

This report was prepared under the Organizational Effectiveness Research Programs of the Office of Naval Research under Contract No. N00014-77-C-0096.

TABLE OF CONTENTS

INTRODUCTION

No exact date can be designated as marking the birth of organizational development. Perhaps the late 1950's or early 1960's marked the first use of the specific term. What has taken place, therefore, has occurred within the last 15 years, years which have seen a substantial investment in the range of activities loosely representing this applied field. Although we lack exact dollar counts, a plausible estimate of the total funds invested in organizational development must run to hundreds of millions of dollars. By any standard, this is a large amount, one that no entity -- whether it be public or private -- may take lightly.

This same time period represents as well the first point at which it was conceptualized as organizational development, as opposed to management development or simply training. No exact definition has general currency, but the term is generally taken to refer collectively to an assortment of training or therapeutic interventions whose aim is presumed to be improvement of the organization and its members.

However it is operationally defined, the problem of organizational development and change would appear to contain two component subprocesses, diagnosis and therapeutic intervention (Bowers & Franklin, 1977). Although equally crucial to the success of any development effort, diagnosis takes prior importance for the simple reason that it occurs earlier in the flow of developmental events, Thus, an important part of the consultant's role is often presumed to consist of translating a wide variety of symptoms into a coherent pattern that permits planning and carrying out appropriate

remedial action. According to Lawrence & Lorsch (1969), the reasons for the importance of diagnosis in organizational development are many and persuasive:

(1) The client system may not be aware of the problem at all. For example, the difference between present effectiveness levels and unanticipated opportunities, rather than obvious difficulties, may be the "problem."

(2) The client system may not be aware of the <u>real</u> problem.

(3) A discrepancy between actual and desired outcomes does not explain and account for itself.

(4) Problem variance is likely to be multiply caused.

(5) Causes are likely to interrelate in complex ways.

(6) Causes are likely to differ greatly in potency, and what is desired is a designation of variables with <u>leverage</u>.

(7) Meaning can only be given to causal information by casting it into an appropriate configuration against a set of principles.

(8) What is required for action planning is an overall and integrated view, not a parochial one.

(9) Diagnosis, if done well, provides some insurance against rushing into an inappropriate treatment that may prove damaging.

In contrast to this, an article by one of the present authors (Bowers, 1976) turned attention to assumptions concerning the consultant's diagnostic role in O.D. The points made there bear repeating.

- While a number of writers have attributed a diagnostic role to consultants, what goes unrecognized is that their diagnoses are often put to little other than heuristic use (that is, they are used merely to stimulate an interesting discussion).
- An unpublished study of consultants' diagnostic skills showed (a) inability to agree with diagnostic conclusions more formally obtained, and (b) more positive change occurring where consultants did relatively little diagnosing than where they did a great deal of it.
- Most consultants currently employ diagnostic methods which rely upon one observer--the consultant himself or herself--to obtain data. The N is restricted, not only in this fashion, but also by the fact that this consultant-observer is limited to a time-bound behavior sample.

These observations should not surprise us. Findings from the general field of assessment and classification have provided strong support to the position that statistical prediction is superior to non-statistical or judgmental methods (Cronbach, 1960). For example, in Meehl's (1954) major review of clinical versus statistical prediction, it was found that statistical prediction was equal to or superior to clinical prediction in 19 out of 20 cases.

Citing this body of accumulated evidence, Cronbach explores the reasons for perenially poor showings by (clinical) judges:

- Judges combine data by means of intuitive weightings which they have not checked.
- Judges casually change weights from one case to the next.
- Judges are unreliable, in the sense that the same case might not be judged the same way twice in succession.
- Judges have stereotypes and prejudices which affect their judgments.

His conclusions are the following:

> "What does this imply? It implies that counselors, personnel managers, and clinical psychologists should use formal statistical procedures wherever possible to find the best combining formula and the true expectancies for their own situation. They should then be extremely cautious in departing from the recommendations arrived at on the basis of the statistics..." (p. 348)

If this is the desirable state for organizational development as well, it is scarcely what in fact obtains. Levinson (1972, 1973), in his published remarks which led to the celebrated exchange with Burke (1973) and Sashkin (1973), stated that there is little resembling formal diagnosis in O.D. Consistent with Kahn's (1974) observations, Levinson stated that the field is characterized by "ad hoc problem-solving efforts and a heavy emphasis on expedient techniques." Tichy (1974) does not reassure us when he finds, in his systematic empirical study, that change agents (consultants) seem to have limited diagnostic perspectives, that their

diagnostic frameworks are rather closely limited to their personal values and goals, and that the potential for intrusion of bias is not small.

Unfortunately, recommended alternatives are relatively scarce. Levinson's recommendations build upon a view and a method of organizational diagnosis that is an extension of the clinical case method. While large amounts of empirical data would be gathered, injecting a clinical judge between the data and the conclusion runs the risks listed above by Cronbach.

On the other hand, this is not the situation nor the age for "raw" empiricism. As the lengthy discussions nationally about discrimination in testing have revealed, in the interest of fairness and equal treatment, more must be taken into account in a decision process than any simple set of numbers, especially where connections between the numbers and real world events may not be obvious. In a similar vein, the sudden rise of the assessment center concept has shown that an appropriate criterion in this day and age (in employee selection, but by extension to the problem of treatment selection in O.D.) must include demonstrable connection between the measures used and the operations or functions performed in the real organization.

These facts lead us to the following preliminary conclusions, which form a starting point for the research to be undertaken in the present report:

- The base of scientific knowledge which undergirds organizational development, while it is growing rapidly, is still remarkably small.

- Much of what is done is based upon consultants' predilections or fads, not upon solid reasons diagnostically generated.
- There is as yet little that could really be termed rigorous diagnosis practiced within the O.D. profession.
- Here, as elsewhere, statistical prediction is likely to prove far more accurate than clinical, or clinically mediated, prediction.
- Raw empiricism, in the form of predictors not obviously related to the processes and functions being diagnosed, no matter how seemingly accurate, are no longer societally acceptable. Prediction must be based upon measures derivable from solid scientific evidence about organizational functioning.

To understand what is or must be involved in diagnosis, we turn to a field which has practiced and taught diagnosis for years and decades, or even centuries: medicine. Ledley and Lusted (1959), in what must be counted as a seminal article, dealt at some length with the reasoning foundations of medical diagnosis. Exhibit 1 presents a few of the principal points which they make, along with organizational diagnostic analogs. In the next sections we present a brief discussion of the content of each point.

EXHIBIT 1

ISSUES IN DIAGNOSTIC REASONING

Medical Diagnostic Issues	Organizational Diagnostic Analogs
1. Symptom complexes (patterns of symptoms) are compared to disease complexes.	1. Patterns of actual organizational characteristics are compared to normative patterns of organizational characteristics.
2. Diagnoses are probability statements, not statements of certainty.	2. Diagnoses are probability statements, not statements of certainty.
3. Diagnosis aids the physician in selecting an optimum treatment under the ethical, social, economic, and moral constraints of our society.	3. Diagnosis permits the selection of an *optimum treatment* (intervention), given society's ethical, social, economic, and moral constraints.
4. The function of the knowledge base is to reduce the logical basis from all conceivable combinations of disease-symptom complexes to only those that actually occur.	4. The accumulated organizational knowledge base reduces available data to a manageable list of potential patterns of characteristics.
5. Maximizing the number of persons cured is equivalent to maximizing the probability that the individual patient will be cured.	5. Maximizing the number of units showing improvement is equivalent to maximizing the probability that an individual unit will show improvement.

Symptoms and Disorders

The total pool of available characteristics (of client units) is at any time limited to those which our knowledge base contains some information about and which our measurement methods are capable of measuring. All available characteristics are, at some level on their respective scales, potential symptoms. Whether they are, in fact, regarded as "symptoms" or not depends upon what past research and experience has found to be true -- that is, what has been added to the knowledge base.

What, then, are diseases, disorders, or states of organizational dysfunction? A disease is a hypothetical construct -- a theoretical term used for convenience purposes to refer to a whole chain of physical events which are hypothesized as having occurred. "Proof" that the hypothesized sequence has occurred (or is occurring) is obtained by some form of validation process. This validation can be concurrent or even retrospective: if little Johnny has influenza, he should display particular additional characteristics or should have displayed them within the last 24 to 48 hours. It can also take the form of construct validation, that is, of showing that only those observables that are hypothesized as going together in fact appear. Finally, the validation process can be predictive: we can wait to see whether subsequent, predicted signs of influenza appear in little Johnny's case. Throughout this sequence of comparisons, however, "influenza" is a hypothetical sequence of events which we presume to be able to see specific signs of at specific points in time. Its excellence as a classification category at any given point in the profession's development is entirely dependent upon the quality and completeness of the knowledge base from which we work, as it relates to the distinctions between this category and others.

What, then, determines what a disease is? It is the generalization and codification processes which past knowledge generators have gone through in integrating the findings from research and experience. Diagnostic procedures which rely upon "expert" assignment to diagnostic categories simply substitute the expert clinician for more public and replicatable listings. If the experts' procedures are unreliable, their classification is, as a criterion, worthless. If they are reliable and valid, it is a valuable aid -- a shortcut to employing the knowledge base directly and in its entirety.

Regardless of the way in which we mediate the process by which the knowledge base's contents get represented, the disease, disorder, or dysfunction is nothing other than a string of symptoms very much like those which we look at in any particular case. It is to this hypothetical symptom string or pattern that comparison is made in a diagnosis.

Diagnosis as Probability Statements

In organizational development and change, the diagnostic process follows essentially this same pattern. Symptoms are organizational characteristics which past research indicates go together to define some more general statement of organizational health or dysfunction. That our "diseases" do not have exotic names in Latin should not dismay us. Perhaps the absence of names at all is an advantage, in fact. Certainly there have been fewer years and resources available as yet for the codification of the knowledge base, and our professional schools teach us to be hesitant, cautious, and qualifying in our statements, rather than authoritative, definitive, and final. These are issues of style, however, rather than substance. The fact remains that there is an

existing knowledge base, comparison to which permits us to make a probability statement concerning any case at hand. Here, as in medicine, a diagnostic statement is a "best guess."

Relevance to Treatment Selection

The whole purpose of a diagnosis is to permit the selection of an optimum treatment or intervention. Here, as in the case of medicine, such choices are subject to social, ethical, economic, and moral constraints imposed by the society in which we live. Certain interventions may be socially unacceptable or even morally offensive. For example, intensely confrontational techniques are clearly unacceptable in many more traditional organizational settings, and under certain circumstances it is conceivable that top management team development training might generate an in-group clubbiness whose effects are racially discriminatory and therefore morally offensive. Other interventions, no matter how appropriate and promising, might be so expensive as to be prohibitive, while still others that would solve the problem might lead to violations of privacy and confidentiality which must be judged to be unethical.

However, within the limits which these constraints impose, the problem becomes one of selecting an optimum treatment from a pool of those available. What is optimum? Ledley and Lusted (1959) turn to value (decision) theory in an attempt to answer this question. Bowers and Hausser (1977) have shown how the organizational development problem can itself be cast into these same terms, and have presented empirical evidence about a limited number of intervention strategies.

A diagnostic procedure which clearly differentiates cases to which each of the known and available interventions are appropriate would obviously be superior to one which, in some measure or other, was unable to distinguish a condition calling for one intervention from a condition calling for another. At the most undesirable extreme would be a "diagnostic" procedure whose conclusions lead always to the same treatment or intervention, a condition which Levinson (1972) implied occurs in organizational development all too frequently.

Role of the Knowledge Base

Even with a relatively simple rating system of "Yes-No" or "Present-Absent," a list of N possible characteristics produces 2^N potential combinations. The number of potential "diseases" or dysfunctional states -- represented by the number of cells in an N-dimensional lattice -- is obviously unrealistically large. In any comprehensive scheme, all of the available units in the world probably would be insufficient to providing a single case per cell. The equally obvious conclusion is that most cells are empty, that they represent nonexistent disorders, and that only a relative few comprise the set of "real" possibilities. It is the task of the knowledge base to provide us with current, accurate information about what these possibilities are.

Much the same point is made in the theory of adaptation in natural and artificial systems (Holland, 1975). Combinations of characteristics rapidly generate astronomical numbers of possibly adaptive structures. If the organism or system were to choose an enumerative adaptive plan -- simply running down the list randomly

until it found the one that worked -- adaptation would rapidly become impossible. As the writer just cited indicates, given even the fastest computers in existence, it would require "a time vastly exceeding the age of the universe to test 10^{100} structures." Instead, adaptive plans to be feasible must be robust -- that is, they must be efficient over the range of situations which will be encountered. One general requirement, therefore, is that an adaptive plan must retain advances already made, as well as parts of the history of what has occurred. This information, of course, is what constitutes the knowledge base in any diagnostic system.

Improvement Probabilities and the Single Case

At first blush, the statement seems unfeeling or insensitive that we maximize the probability of any individual unit's showing improvement when we apply to it a strategy shown to maximize the number of units showing improvement. Organizational development is, after all, a human practice profession, and it seems impossible to ignore facts obviously at hand (within range of our personal observations, for example).

Nevertheless, observations based upon an N of one (observer), collected under atypical conditions and within nonrepresentatively short time frames, are no more reliable and accurate when taken singly than they would be if used en masse for large numbers of cases.

This issue was touched upon by one of the present writers in an earlier report: "Even the most accurate diagnosis may suffer from mid-stream or horseback revisions made by the consultant as he approaches its use. Basically, any data collection and analysis method treats with some

degree of care and accuracy a portion, but not all, of the behaviors, events, and issues in the life space of the client system. Some portion is unique to that system, or to any group within it, or will have been excluded from the array of information categories designed in the diagnostic process at its inception. As the consultant approaches a particular unit or group of the client system, he will necessarily see other aspects of what he feels are its functioning not represented in the diagnosis which he has in hand. Since he is dealing with a real client, in a real world situation, the temptation is well nigh irresistable to revise the diagnosis on the basis of his current observation. Yet, he is one observer observing at best a limited and time-bound behavior sample. To the extent that he makes such revisions he therefore very likely reduces both the reliability and the validity of the diagnosis with which he works. Said otherwise, he approaches each group, or each setting, as a unique instance with live people and real problems. Yet in many ways the diagnosis and treatment problem in organizational development is a "large N" problem. Were he to work on the basis of the diagnostic data provided to him and that alone, given that it is reliable and valid, he would, across a large number of cases, succeed in a high portion (assuming that the diagnostic and prescription processes are themselves high in quality, reliable, and valid). Yet he does not ordinarily approach his role with that degree of objective detachment, and each time that he yields to the temptation to revise on the basis of "current reality" he submits himself to a situation in which his action steps are based on less than acceptably reliable and valid data." (Bowers, 1974)

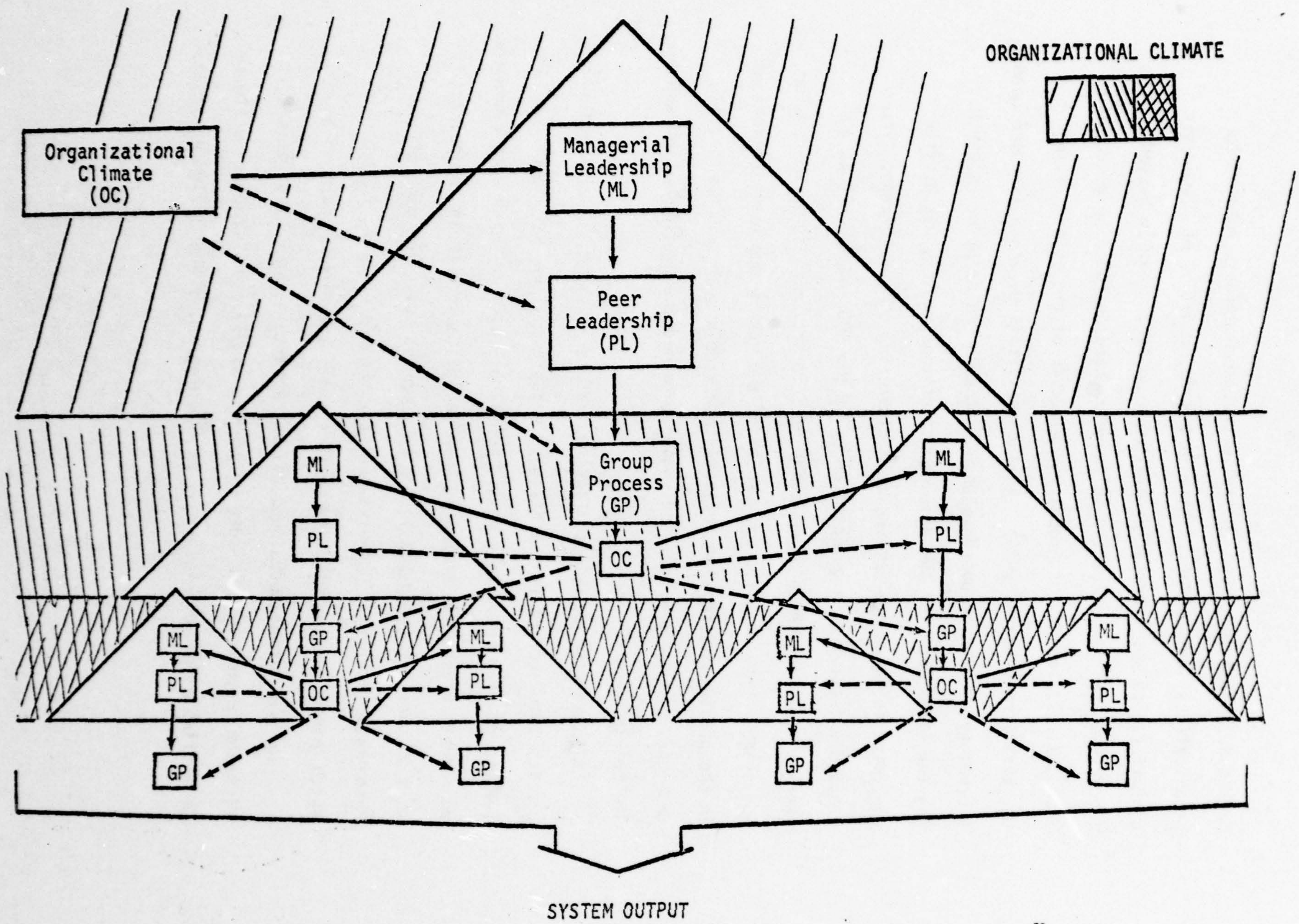
Toward Relevant Research

Clearly, therefore, any attempt to develop and test more rigorous diagnostic procedures in organizational development should be based upon a model containing principles of organizational functioning. In other words, it should be theoretically anchored to a conceptual statement that is itself both organizational in content and comprehensive in scope. While the literature on organizational management is ripe with theoretical statements, most of them do not meet criteria of acceptability for our present purposes. Many must be dismissed as less comprehensive than is necessary for the present problem: that is, they are elegant treatments of an isolated issue such as job design or individual satisfaction or leadership, but they ignore other areas. Others may be rejected because, although they encompass most of the domain, they are lacking in adequate empirical underpinnings. However, one theoretical statement which does appear to meet the criteria just outlined is the Likert meta-theoretical paradigm (Likert, 1961, 1967, 1976; Bowers, 1976). It is this theoretical statement which underlies the measures collected in the data bank to be used in the research represented in this project.

Most recent evidence suggests that this paradigm assumes the form taken in Figure 1 (Bowers & Franklin, 1977). As a set of principles, this paradigm would appear to satisfy the criteria of comprehensive and evidential validity (Bowers & Franklin, 1977; Likert, 1977). It is operationalized here in the form of the Survey of Organizations, a machine-scored standardized questionnaire which has been used in various editions since 1966 to collect organizational survey data for assessment, feedback, and benchmark purposes (Taylor & Bowers, 1972). Portions of these

Figure 1

ORGANIZATIONAL FUNCTIONING



banked data have been used in earlier research efforts related to organizational development. In this regard, a method of diagnostic classification was previously developed and preliminarily tested. Termed CANOPUS, it contains a software package designed to generate a diagnostic statement for groups and pyramids of groups comprising organizations (Bowers, 1974). This classification method is based upon a typology of work groups developed in the course of prior research. The technique used for the development of the typology was profile analysis, in which one arrives at a clustering of work groups. The profile consisted of a group's scores on the SOO indexes and as a profile reflected three basic kinds of information: level, dispersion, and shape. Level was the mean score of the work group over the indexes in the profile; dispersion reflected how widely scores in the profile diverged from the average; and shape concerned the profile's high and low points.

A measure of profile similarity that takes shape, level and dispersion into account is the distance measure. If one considers a person (or group) as a point in a multidimensional space in which each dimension represents a variable or index, then the distance between two points, that is, persons (or groups), can be computed using the generalized Pythagorean theorem. The distances can then be examined to determine which groups cluster together in that multidimensional space.

The clustering technique, called hierarchical grouping, uses this distance measure as a measure of profile similarity. Computer software is available for this technique in the program, HGROUP (Veldman, 1967).

This program begins by considering each original object, in this case a work group, of those to be clustered, as a cluster. These N clusters are then reduced in number by a series of step decisions until all N objects have been classified into one or the other of two clusters. At each step the number of clusters is reduced by one through combining some pair of clusters. The particular pair to be combined at any step is determined by the computer's examining all the available combinations and choosing the one which minimally increases the total variance within clusters. It is this latter minimizing function that utilizes the distance notion. The total variance within clusters is a measure of the closeness of the points in multivariate space in clusters already chosen. A substantial increase in this variance, which the HGROUP program labels an error term, indicates that the previcus number of clusters is probably optimal for the original set of objects or work groups. The program provides an identification of those groups contained in each cluster so that further analyses can be conducted on phenomena within clusters.

The HGROUP program was applied to three random subsamples drawn from the data bank (Hausser & Bowers, 1977). When the three sets of data were considered jointly, a total of 17 distinct profiles emerged. In many ways, this software system would appear to meet generally the requirements listed:

- It compares data to appropriate norms.
- Problems once identified are prioritized in terms of their potential impact upon outcomes.
- It seeks causes for observed conditions among situational, information, skill and values conflict predictors, employing a <u>distance statistic</u>.

. It selects a broad set of potentially appropriate action steps from an array of possibilities.

It converts the whole and its parts into a readable narrative by computerized text-writing.

Still, the outcome of the method is based upon the measures, and those measures derive from the theoretical paradigm previously cited. While attractive, it is but one of several statements that might have formed the basis for operations and measures. Clearly some difference among theorists is to be expected. The domain is sliced differently, and the terms applied to collections or clusters of behaviors and processes will vary substantially. However, if the fundamental, general algorithm is the same, we can at least be somewhat reassured that subsequent work will not be unacceptably parochial.

In an effort to address this question, in an earlier report the writings of nearly 30 prominent persons in the organizational management field were examined. (Bowers, Davenport & Wheeler, 1977) The conclusion from this examination of the field was that we can be reasonably confident that a common algorithm underlies most of the major works in it. It leaves us reassured that adherence to an alternative formulation from this list, if pursued to its most basic form, would not result in an utterly different diagnostic scheme. Terms might be different, and the operations employed by each writer to measure particular sets of variables might vary widely, but the rationale and the set of primitive constructs would be very much the same.

A Stock-Taking and Some Implications

Against the expressed need for improved methods for diagnosis in organizational development we can array the following major points from the preceding discussion. An adequate diagnostic procedure necessitates:

(1) A theoretical model which is acceptably comprehensive, which shares the same general algorithm present in the array of principal alternative formulations.

(2) A bank of data, collected by a standardized instrument in a wide variety of organizational settings, both military and civilian.

(3) A recognition that accuracy in diagnosis will, here as elsewhere, very likely be enhanced by statistical operations rather than clinical judgment.

(4) An acknowledgment of societal requirements rejecting "raw" empiricism in favor of statistical procedures and measures which are content valid.

Considering the magnitude of the problem, the size of data sets, and the turnaround time requirements present in most organization development situations, yet another requirement would appear to be present: that whatever operations result be computer-assisted. In this area, one can profit from the experience of another practice-oriented profession whose researchers have explored computerized diagnostic procedures, psychiatry.

There, as elsewhere, substantial difference of opinion exists concerning the best method of evaluating the importance of symptoms. Two general types of models relying upon probability statistics have been proposed:

A. A discriminant function model, in which each symptom is given an empirically derived weight, and an artificial measure is then obtained as the sum of the weighted values (Crooks, Murray & Wayne, 1959).

The arguments in favor of a multiple discriminant model are at least threefold:

(1) It better replicates the thought process employed by the human diagnostician, who does not treat each symptom in present/absent fashion, but rather attaches greater or less weight to each symptom according to past experience (i.e., his/her own version of the knowledge base.)

(2) Symptoms which correlate highly with the presence/absence of a disease are given more weight than those that have shown little or no correlation to its presence/absence.

(3) Appropriate weighting also depends upon a symptom's correlation with other symptoms. If the overlap is high, then one would weight the second symptom much lower than would be the case if the two symptoms have little relationship to each other.

At least two objections have been raised to this method:

(1) It relies upon the accumulation of a large developmental sample of cases, which is difficult, expensive, and in most instances unlikely.

(2) It capitalizes upon accidental features of the developmental sample and thus gives an inflated estimate of its accuracy. If the validation sample comes from a somewhat different population, the drop in efficacy is even greater.

B. A Bayesian or frequency-count model, in which the relative frequency of occurrence of each possible symptom-disease pattern is considered (Ledley & Lusted, 1959).

The principal argument raised in favor of a Bayesian approach to computerized diagnosis appears to be that it also is claimed to model the human judgment process by which symptoms are converted into a diagnostic statement (that is, that the physician, for example, employs a conditional probability judgment process in arriving at a diagnosis.)

The objections are a bit more extensive:

(1) It is difficult, if not impossible, in diagnostic work to satisfy the conditional independence requirement (the requirement that the probability of finding one particular symptom given that the disease is present, is unaffected by the presence or absence of any other symptom.)

(2) As in the other statistical method, it requires the accumulation of a large developmental sample of appropriate form and content.

(3) The necessary assumption that the diseases are mutually exclusive may not hold.

(4) As in the other statistical method, it capitalizes upon accidental features of the developmental sample.

To these have been added a third method which:

C. Treats the issue as a decision-tree problem, thus relying maximally upon excellence of the knowledge base and not at all upon probabilities in a developmental sample (Spitzer & Endicott, 1968).

The arguments in favor of this method are given by at least one proponent (Spitzer, et al, 1974) as the following:

(1) It is independent of any specific body of data; that is, it does not require a large developmental sample.

(2) It is not constructed so as to be optimal for any one population and for this reason "travels well" from one setting to another.

(3) As in the case of each of the other two methods, it is thought to represent optimally the thought processes of the human diagnostician.

The objections are the following:

(1) It is quite dependent upon the accuracy of the theory which underlies the decision tree and is therefore ultimately as dependent as the other methods upon past data accumulations, their care and form.

(2) Its generalizability may be more apparent than real.

(3) Its assumption that the diseases are mutually exclusive may not hold.

Comparison of the Three Methods

Several efforts have been undertaken in psychiatry to compare two or more of these methods empirically. The results are best described as decidedly unclear. Overall and Hollister (1964) conducted one such comparison, but, unfortunately, the rules used by their computer programs were obtained from diagnostic stereotypes provided by experts, rather than from observed characteristics of actual cases.

Melrose, et al. (1970) compared a multiple discriminant with a decision-tree approach and found that: (a) on single assignments the decision-tree approach showed a greater degree of agreement with an expert judgment criterion; (b) if first, second, or third possible assignments were allowed, the multiple discriminant method showed a greater degree of agreement than did the decision-tree; and (c) in any event, each method performed better for certain diagnostic categories.

Finally, Fleiss, et al. (1972) compared all three methods and found none of the three to be clearly superior to the other two. Again, however, the criterion was agreement with expert diagnoses, a criterion whose unreliability the authors duly note.

A Dilemma and Some Issues

The work from psychiatry, just cited, contains a dilemma whose existence questions the whole body of findings and whose resolution might be seen as rendering the whole exercise rather trivial. Elegant, replicatable, and readily transportable methods are designed and tested against a criterion of "judgment by expert clinicians." Yet here, as elsewhere, the

Cronbach warnings apply: expert clinical judgment is notoriously unreliable. What has been developed, therefore, are three elegant ways of replicating an unreliable procedure. On the other hand, had a reliable, replicatable, transportable procedure existed for use as a criterion, it would no doubt have been more sensible to employ it as the diagnostic method, rather than as a criterion for other methods.

Several implications stem from this observation. First, where in psychiatry expert clinical judgment is an unreliable classification method and criterion, in the present instance we do possess a reliable, verifiable procedure, one based upon the distance statistic. An issue of some importance lies in these facts, that a distance function was used to generate the original typology and that at least the discriminant function bears a close kinship to that method -- closer than does the decision tree procedure, for example. This breaks down into two component questions: (a) to what extent is the typology an artifact of the method, and (b) to what extent are our findings simply evidence of the fact that nothing can approximate a distance function like a distance function?

To address this, let us suppose that the total population of approximately 7,000 groups were uniformly distributed in the 14 space represented by the measures. To generate the typology, we originally selected three samples totalling 533 groups. We found that the clustering procedure classified these into 17 clusters. Was that clustering a chance outcome, a "forcing" into clusters? The original results suggest that the probability of that was extremely low. Still, let us suppose that it was that once in 50 or 100 times when such a result would occur by chance. What would then be the result of attempting to fit the entire population to the types

so generated? Obviously, we would not expect a fit to occur -- by a distance function or any other function.

Alternatively, one may approach the problem from a conceptual, rather than an empirical, viewpoint. Let us assume for the moment that the typology is real -- that these really do exist 17 types which have true values on each of the 14 dimensions. In the original study we identified that they do exist and then estimated their values by calculating the mean index scores for the defining clusters. Our estimates may be in some measure discrepant from those true values. Regardless of the fact that a distance function was used to define the clusters, the profile estimates differ from the true values on one or more of the three constituent elements of distance: level, dispersion, and shape. In other words, our estimate of the type's profile value may differ from the true value because it is too high or too low (level), because the component indexes are too compact or too spread out (dispersion), or because the rank order of indexes in the profile is out of order (shape.)

Pulling these alternative lines of thought together, if a typology does truly exist, and we have closely estimated the profile values involved, it matters little that we did so by a method that clusters according to a distance function. The profiles are for all intents and purposes those of the true types. We can then judge the accuracy of the distance function -- or the discriminant function or the decision tree method -- by the tightness with which that method is able to align the population to the "true" types. In part, therefore, we do two things simultaneously when we classify all cases in the population by a distance function:

(a) we assess the plausibility of the original typology itself, and

(b) if the typology proves to be plausible, we assess the degree (compared to other methods) to which a distance function itself is effective in classifying cases.

However, above and beyond this issue, there are other issues of equal or greater importance, the answers to which are in no way obvious. These issues of evaluating classification procedures include:

- Accuracy Based Criteria
  - Proportion of correct classification
  - Typology reproduction
  - Weighted dimensional distances
  - Zero-one counts
  - Severity of misclassification
- Non-Accuracy Based Criteria
  - Information required to make a decision
  - Amount of data required to develop the diagnostic process
  - Ease of calculation

Of these, one -- weighted dimensional distances -- seemed, and in fact proved, to be beyond the scope and resources of the present project. The remaining seven issues were investigated. A brief description of them is presented in the following sections.

Accuracy Based Criteria

The purpose of this section is to present four different criteria to use in evaluating classification procedures, each of which is based on a measure of accuracy. In later sections, non-accuracy based criteria are discussed.

The four different classification procedures to be investigated are a decision tree (DT), multiple discriminant function (MDF), a Bayes rule (B), and a straight (least squares) distance function (DF). The typology into which work groups are to be assigned has been discussed in earlier reports (Bowers, Davenport & Wheeler, 1977). Two aspects of the typology and its development are relevant to the investigation here. First, it has been derived by a clustering algorithm which, in effect, groups the observations (vectors of mean scores from work groups) so as to minimize the variance within clusters. The variance metric is also a distance metric, which will result in the same classification by the DF for a particular work group as you would get by including that work group in the clustering process originally. The consequence is that we will use the distance function (DF) classification as the correct classification for any particular work group.

The typology developed by the clustering algorithm allows the creation of a vector of scores from the averages across all work groups within the cluster (Bowers & Hausser, 1975). Thus, we can think of the typology containing 17 types as a set of 17 vectors. That is, a type is represented by a single vector of scores.

Proportion of Correct Classification

The classical criterion for evaluating classification schemes is the proportion of agreement with some external "expert" opinion. In the present situation, as discussed above, the expert opinion will be that provided by the distance function (DF). However, there is some interest in seeing how well one data analytic technique is able to reproduce another data analytic technique when additional criteria (beyond proportion of agreement) are considered. The reader is referred to the following sections for discussion of other criteria.

As a beginning, we propose to analyze the three classification procedures, MDF, B, and DT, by comparing the proportions of agreement with DF. The straightforward process will be as follows:

> For each of MDF, B, and DT, compute the proportion of the (approximately) 7,000 work groups which are assigned to the same type by both the given procedure and DF.

The figures obtained from the above analysis can be viewed, for each procedure, as the percentage correct within each type of the typology as well as aggregated across all types. Such a review may enhance the application of the technique in other settings.

Reproduce the Typology

Assume a particular classification procedure has classified, say, k groups into a single type within the typology. It then would be possible to compute the averages for those k groups, resulting in a single vector of index scores. It is desirable to have this vector of scores for the

assigned groups be close to the vector of scores which represent this type. A measure of accuracy of prediction would be to compute the sum of the distances between each of the 17 types and their corresponding mean vectors from the groups assigned to them.

A criterion for evaluating the different classification schemes, then, would compare the sums of distances between the typology vectors and the group means. Because of the use of a standard distance measure to compare the type vector and the vector of average scores from the observations assigned to that type, the distance function will do best here, as well. However, it would be appropriate to compute the sum of the distances for DF to use as a standard against others which could be compared.

The procedure to implement this process, using all work groups, is as follows:

For each of MDF, B, DT, and DF:

1. Compute a vector of average values for each set assigned to a type by the procedure.
2. Compute the distance between this vector of averages and the type vector.
3. Sum the distances across all types.

For each of MDF, B, and DT:

4. Compute the ratio of the sum to the sum for DF.

Zero-One Count

In assigning a work group to a single type within the typology, two different patterns of work group scores may lead to the same distance from the type. Consider the following example, simplified to reflect only four dimensions.

type	(2,2,2,1)	
work group 1	(4,4,4,3)	distance = 4
work group 2	(2,2,2,5)	distance = 4

One approach to overcoming the above situation is to use the weighted dimensional distances described earlier. An alternative approach is to count the number of dimensions for which the assigned group is significantly different from the type. If we define significantly different as being greater than or equal to two in the above example, the counts would be four for work group 1 and one for the second work group. Mathematically, this counting can be represented as the sum of the values c_i,

$$c_i = \begin{cases} 1, \text{ if } |Y_i - X_i| \geq s \\ 0, \text{ if } |Y_i - X_i| < s, \end{cases} \quad \text{and } C = \Sigma c_i$$

Here s is the value defining significantly different, and $\underline{Y}$ and $\underline{X}$ are the vectors representing the points of interest. Then C may be computed for each work group assigned to a particular type, and an average computed for each type. The choice of s should be based on some measure of relative magnitude or theoretical argument. For the present study, we propose s be chosen as a measure of the average variability of the SOO indexes.

Thus, the average of C could serve as a criterion for evaluating the accuracy of classification. The procedure involved would be:

(1) Assign a value to s, the significant difference.

(2) For each procedure, compute C for each work group.

(3) Compute CT for each procedure.

Severity of Misclassification

One argument against the use of the classical frequency of correct classifications is that it treats all misclassifications as equal. That is, there is no distinction between misclassifying a particular work group into any of the k-1 incorrect types in a typology of k types. It is not unusual, however, for the cost of misclassification to be widely different across the k-1 incorrect types, as well as being contingent on the correct type. For example, incorrectly diagnosing an individual with a severely sprained ankle as having a broken ankle or having leukemia has different costs; additionally, incorrectly diagnosing an encephalitis case as a broken ankle or leukemia has yet different costs.

An alternative to the straight proportion of correct classification criterion of accuracy is one which allows for differential costs of misclassification. In particular, we might consider variations upon either of two different costing models for misclassification:

1. That a cost of 0 be assigned to any misclassification of a work group into a category which calls for (a) treatment if so does the correct category, or (b) non-treatment if so does the correct category; otherwise the cost is 1.
2. That a cost of 0 be assigned to any misclassification of work group into a category which calls for a treatment which is known to have a positive effect on the correct category, and a cost of 1 if the incorrectly selected category calls for a treatment which is known to have no effect, or a negative effect, on the correct category.

These are only two models of a vast number of possibilities for costing misclassification. However, they are reflective of the primary concerns of misclassification -- the application of inappropriate or harmful treatments. In the present instance, we have chosen to use a variation upon the second form of costing -- one which takes both direction and severity into account.

Bowers and Hausser (1977) examined the effects of different change strategies on each of the 17 groups of the typology. They rated the effects either as negative, neutral, or positive. For number 1 above, it is possible to define a work group calling for treatment. Otherwise, the work group calls for no treatment. The results of Bowers and Hausser (1977) can also be used to establish the appropriate pattern of costs for number 2 above.

Non-Accuracy Based Criteria

Information Required to Make a Decision

One criterion by which one can compare diagnostic techniques is the information required to reach a diagnosis. If two techniques perform equally well when the full set of indexes is used in the diagnosis, then it will be to the diagnostician's advantage to use the one which requires less information to reach a decision. Collecting and processing information is costly, in terms of money, time, and complexity of processing. For instance, if one can obtain results using three pieces of information that are as good as the results using five pieces of information, it will clearly be advantageous to use only three pieces of information.

The most obvious way of reducing the amount of required information is to find an appropriate way to reduce the number of indexes that are used in the diagnostic process. There are two ways of reducing the number of indexes. The first is to discard or eliminate indexes that have been

shown to be unnecessary for the diagnosis. The second is to combine several indexes into "super-indexes." For instance, it might be possible to combine the four peer leadership indexes into a single index.

Another method of reducing the information required to make a decision is to reduce the number of respondents proportionately from an organizational survey. To investigate this approach, we propose to draw two random samples (67% and 33%) of individuals from the total data set and classify the "reduced" work groups. Evaluation of the effect of this reduction will be the proportion of agreement in classification between the use of reduced data and full data.

Amount of Data Required to Develop the Diagnostic Process

Of interest to the researcher is not only the amount of data that must be processed to obtain a diagnosis, but also the amount of data required to generate accurate diagnostic processes.

All of the techniques to be tested require some kind of historical data base, from which the diagnostic process is generated. In all instances, larger data bases should provide more accuracy than smaller ones. If one technique can be generated quite accurately from a smaller data base than another, that technique would be preferable, since it would be more cost effective.

The amount of data necessary for the techniques will be observed variously, ranging from sample sizes necessary to stabilize estimates to the generalizability of procedures computed from a developmental subsample to the remaining sample.

Ease of Calculation

The ideal diagnosis technique is not only accurate, but relatively simple to perform. In the organizational diagnosis situation, the diagnostician may very well be a change agent who must make treatment recommendations while on site or otherwise out of contact with computer facilities.

Three ease-of-calculation factors are:

1. Can be calculated on-site versus in a central location.
2. Can be done with a hand calculator or by hand versus requiring EDP facilities.
3. Few things to be calculated versus many things to be calculated.

While not identical, these three factors are obviously not completely independent of one another.

The comparison of diagnostic techniques on these factors will be a matter of subjective judgment. It is proposed that after the optimal solutions have been determined for each diagnosis technique, the researchers create a table indicating their assessment of each technique on the ease of calculation factors. The diagnostician will then be able to select among the techniques when ease of calculation is an important aspect of a project.

A Restatement of the Research Questions

From the concepts and discussion presented in the preceding pages, we propose to examine findings germane to the following questions:

(1) Is there evidence that the original typology holds up when applied to the entire population of groups; that is, does it appear to reflect reality?

(2) If the typology holds up, how effective is a distance function in classifying by type?

(3) Using the distance function classification as the criterion, how accurate are discriminant function, decision tree, and Bayesian methods in classifying groups?

3a. What proportion of cases does each method classify correctly? (Proportion of correct classification)

3b. How well does each method do in reproducing the typology; that is, how close is the vector of scores for assigned groups to the vector of scores which represent the type? (Typology reproduction)

3c. How do the methods differ in number of dimensions and which scores are significantly different from the type values? (Zero-one count)

3d. How do the methods differ in the cost lines (in change outcomes) of misclassifying groups? (Severity of misclassification)

(4) To what extent are the methods able to classify groups correctly using reduced information sets?

4a. Can data from samples of respondents be used equally well as a complete census of group members in classifying groups?

4b. Can less than all indexes be used equally well as all indexes in classifying groups?

(5) Do the methods differ with respect to the amount of data required to develop the method?

(6) Do the methods differ in ease of calculation?

CLASSIFICATION TECHNIQUES

The original development of the typology used in the present research is reported by Bowers (1975) and Bowers and Hausser (1977). In effect, using a clustering algorithm and three different samples of work groups (two civilian and one military), they found that 17 different types resulted. The analysis was done using the same 14 SOO indexes being used in the present research. The 17 types that comprise the typology can be thought of as 17 points in (14-dimensional) space.

The classification issue is one of assignment of new work groups to these different points (types). Four different procedures historically have been used in a variety of classification applications. They are briefly described below:

(1) Distance function: Find which type a new unit is closest to with respect to some metric (usually the typical Euclidean metric).

(2) Decision tree: By asking a pre-planned series of questions, which, depending on the pattern of responses, leads to a final state or type within the typology.

(3) Discriminant function: From a developmental sample one computes a set of weights to be assigned to indexes, resulting in an equation for each type in the typology. By applying all equations to the score values of a new unit, the unit is classified as belonging to that type for which the resulting equation value is largest.

(4) Bayesian: Based on the development of prior probabilities of different score patterns for different types, posterior probabilities are calculated from the score values of a new unit for each type in the typology. The type for which this posterior probability is greatest is the type to which the new unit is assigned.

Research into the effectiveness of these different procedures for the purposes of medical diagnosis is mixed. As indicated earlier, different studies have claimed success for each of discriminant function, decision tree and Bayesian approaches.

The present research is to investigate the applicability of the above techniques to the field of organizational diagnosis. The purpose of this section is to outline the techniques only as they were applied herein. Later sections will discuss the effectiveness of the various techniques.

Distance Function

The distance function classification procedure first computes the distance between a new work group's index scores and each of the 17 types. The work group is then assigned to the type to which it is closest. Notationally, let $(x_1, x_2, \ldots, x_{14})$ be the index scores for a new work group. Let $(m_{j1}, m_{j2}, \ldots, m_{j14})$ be the vector of values for the jth type in the typology. The (Euclidean) distance from the work group to the jth type is given by

$$D_j = \sqrt{\sum_{i=1}^{14} (x_i - m_{ji})^2}$$

The application of this technique:

- does not involve a developmental sample,
- involves nearly 500 calculations, implying the assistance of some type of computer or other programmable calculating equipment,

. does not require a "correct" classification in its development.

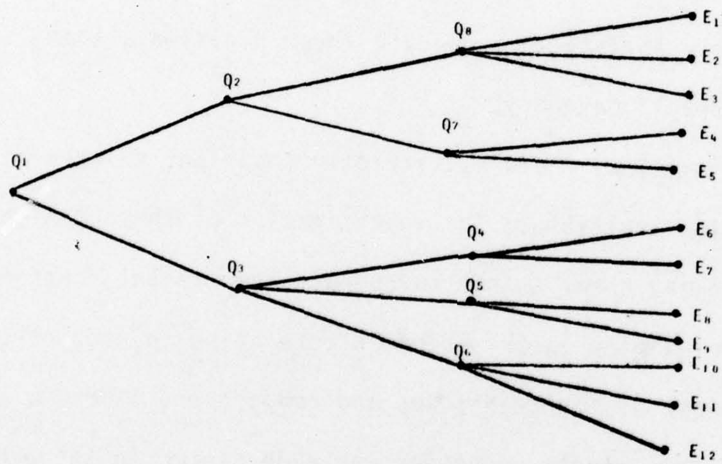
The existence of a "correct" classification is an issue which makes organizational diagnosis, in its present state, different from medical or other diagnosis/classification. In these latter applications, either time (such as in predicting success or failure) or a board of experts (such a group of trained clinicians deciding a patient is schizophrenic) create the "correct" classification. The result is that most numerical techniques are then judged by their ability to replicate these decisions, that is, classify into the "correct" category.

In the present situation, there is, strictly speaking, no external "correct" classification available. The investigation of the numerical classification techniques, then, is not to compare them to the "correct" classification, but to examine their performance relative to each other. The option is to take one of the techniques and compare the others with it. The selection of this single technique was made easier in the present situation by the methodology used to develop the typology. The underlying mathematical approach was to cluster those work groups which were similar to each other, their similarity being judged by their Euclidean distance. Thus, the typology itself, was based on grouping by a distance criterion. Thus, for this study, a new work group's "correct" classification was considered to be that which it was assigned by the distance function.

Decision tree

The approach we have labeled as decision tree is somewhat different operationally from the classical application of decision tree techniques. In the latter, a series of questions is asked sequentially, the answer to one indicating which question is asked next. Pictorially, the tree effect is demonstrated below.

Figure 2



Q_i = Question

E_j = End state

In the present application, a series of questions was asked with the answer to one not affecting which question was to be asked next. In essence, a series of yes/no questions were asked, and the number of favorable answers recorded. The history of this approach is discussed more fully by Wheeler (1978) and Bowers, et al (1977).

In particular, the approach herein can be visually represented as follows:

Figure 3

Index	Type 7 Intervals
1	(2.8309,4.0243)
2	(2.8768,3.9558)
3	(2.2041,3.2587)
4	(2.3448,3.4100)
5	(3.0014,4.0986)
6	(2.4958,3.5316)
7	(2.3023,3.3471)
8	(2.3070,3.3290)
9	(2.1423,3.1649)
10	(2.0692,3.0188)
11	(2.4115,3.3771)
12	(1.9535,2.8609)
13	(2.8355,3.8083)
14	(1.5535,2.6925)

In the above figure, a new work group's values for the 14 indexes are compared to the intervals given for Type 7. The number of intervals which contain the new index values are recorded, and that becomes the fit for Type 7. The fit for each of 17 possible types is computed, and then the maximum of these fit values is ascertained. The classification rule is to place the new work group in that type for which the fit is the largest. The results of this approach are presented in the next chapter.

It is possible, it should be noted, to insert the above form into the classical decision tree model applied to each type. The technique, then, is the application of 17 decision trees. If, in Figure 2, Question 1 (Q_1)

was whether the new work group's first index falls in the appropriate range for the first index, and both Q_2 and Q_3 were to ask if the second index is in the range for the second index, Q_4 through Q_8 would ask the same for the third index, etc. down through the 14th index, then the End state (E_j) would be the fit for that work group for that type.

The last feature of implementing this technique has to do with the selection of the intervals. That is, for each of the 17 types and 14 indexes, intervals were constructed. While a variety of procedures might be used, some of which may not be data based, it was chosen to use symmetric intervals about the ideal value. The distance about the ideal value was taken to be a certain number of standard deviation units. For example, for Type 7, index #1, the ideal value is 3.4276, and 1.5 standard deviation units is .5967, yielding an interval of (2.8309,4.0243) (see Figure 3). The ideal values were taken to be the index means of the groups of each type resulting from the original development of the typology.

To apply this form of the decision tree approach, it

- does require the establishment of intervals, most likely from a developmental sample, which have been "correctly" classified,
- does not require involved mathematical calculations in its assignment of new work groups, thereby permitting either hand or computer use.

Discriminant Function

The discriminant function approach is a data based classification procedure. That is, from a developmental sample, a set of linear (or quadratic) functions of the predictor variables are generated which maximally differentiate among the groups of the typology.

To apply the discriminant function technique, it is assumed that one has a set of predictor variables that are not highly interdependent (multicollinear). This is exacerbated by the fact that this requirement is to be met within the subgroups that represent the various classes of the typology. In the present work, interdependence between the 14 indexes (predictor variables) was sufficiently high that it precluded the calculation of the discriminant functions.

The option of reducing the number of predictors was explored. As may be expected, the highest levels of interdependence were found within the three major categories of indexes: organizational climate, supervisory leadership and peer leadership. The approach taken to reduce the number of indexes was to do a cluster analysis on the 14 indexes, clustering by the degree of correlation. The indexes clustered within the three areas just mentioned. However, for example, without taking all four peer leadership indexes into a cluster, those left out would still correlate quite strongly with the cluster (.7 to .85). Thus the decision was made to form three super-indexes:

- Organizational Climate (Decision Making Practices, Communication Flow, Motivational Conditions, Lower Level Influence, Human Resources Primacy).
- Supervisory Leadership (Support, Work Facilitation, Team Building and Goal Emphasis).
- Peer Leadership (Support, Work Facilitation, Team Building and Goal Emphasis).

Thus, all discriminant function classification is based on the three super-indexes, each being the sum of its component indexes.

The total data base available for this project was divided randomly into two equal subsets. These samples contained over 3,000 work groups each, all of which had previously been assigned a "correct" classification by the distance function algorithm. From the developmental sample, discriminant function weights were calculated. The application of these resulting discriminant functions to the test sample is described in a later section.

The application of this technique

- does require a fairly large developmental sample with use in a typology with numerous types,
- does require prior "correct" classification of the developmental sample,
- does require extensive calculations to develop the discriminant functions,
- does require involved calculations to assign a new work group, indicating the use of a computer.

Bayesian Method

Principally, the Bayesian approach involves the establishment of a set of prior probabilities and likelihood functions, the calculation, for a new work group's actual scores, a set of posterior probabilities, and the assignment of that work group to the type for which the posterior probability is the largest. A more detailed explanation for the consideration of this approach is included in Bowers, et al (1977).

For the present work, the development of the prior probabilities was simply the relative frequencies of each type in the entire data set. As discussed in Bowers, et al (1977), the development of the likelihood functions, which are used in the calculations of the posterior probabilities, is pivotal. The family of distributions used to develop the likelihood functions is critical. The multivariate normal family, while mathematically convenient, has the shortcoming that it permits arbitrarily large or small values for its observations. The values obtained from the work groups are means from five point scales, and thus are bounded. The multivariate normal family is also symmetrical in each dimension. For the same types in the typology which have relative large index means, say near 4.7, and say a standard deviation of .3, it is impossible to be more than one standard deviation above the mean (the maximum score being five), and some values do fall below 4.4, resulting in asymmetrical distributions. For these reasons, the multivariate normal distribution was omitted from further consideration.

Another family of distributions which was investigated was the beta family. This family has density functions of the form

(1) $$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \; y^{\alpha-1}(1-y)^{\beta-1},$$

for $o<y<1$, and $\alpha>o,\beta>o$. The different members of the family (for various combination of α and β), are not necessarily symmetrical, and have bounded range $(0,1)$. A simple conversion of a work group score, say x, by $y = (x-1)/4$ results in values in the appropriate range.

In order to fit a member of the beta family to an empirical distribution, one can estimate the parameters α and β by the following:

(2) $$\alpha = \frac{M}{S^2}[M(1-M)-S^2]$$

(3) $$\beta = \frac{1-M}{S^2}[M(1-M)-S^2]$$

where $M = 1/4(\overline{X}-1)$ and $S^2 = \frac{1}{16}V$, $\overline{X}$ and V being the mean and variance of an index on the five point scale. These equations result from the fact that the mean and variance of the beta distribution are given by

$$\mu = \alpha/(\alpha+\beta),$$

$$\sigma^2 = \alpha\beta/[(\alpha+\beta+1)(\alpha+\beta)^2].$$

The procedure for computing the posterior probabilities for a new work group, given the likelihood functions, is as follows:

#1 From among the profiles of types 1 through 17, select one to serve as a standard, say #3.

#2 Let Q_j represent type j, for j=1,...,17. Then let $P(Q_j)$ be the relative frequency of type j (based on the distance function classification of all work groups in the sample).

#3 Let a new work group have a vector of converted index scores given by $I = (y_1, y_2, \ldots, y_{14})$, where $y_i = [(X_i-1)/4$, X_i being the i^{th} actual index score.

#4 Now compute, for j=1,...,17, the value of the odds-likelihood ratio

$$\frac{P(Q_j|I)}{P(Q_3|I)} = \left[\prod_{i=1}^{14} \frac{f_{ij}(y_i)}{f_{i3}(y_i)}\right] \frac{P(Q_j)}{P(Q_3)} \tag{4}$$

where f_{ij} is the estimated density function, given in (1), for the i^{th} index within the j^{th} type. Note that α and β also change for each i and j. Implicit at this point is the prior calculation of the estimated parameters α and β for all 14 (indexes) x 17 (types) cases.

#5 Now compute $P(Q_3|I)$ by

$$P(Q_3|I) = \left[\sum_{j=1}^{17} \frac{P(Q_j|I)}{P(Q_3|I)}\right]$$

#6 Returning to equation (4), it is now possible to compute the posterior probabilities $P(Q_j|I)$ by

(5) $$P(Q_j|I) = \left[\frac{P(Q_j|I)}{P(Q_3|I)}\right] P(Q_3|I).$$

Again note that the selection of profile 3 was arbitrary, and any of the 17 types could have been selected.

#7 Find the value of j for which the posterior probability, $P(Q_j|I)$, $j=1,2,\ldots,17$, is largest and assign the new work group to that type.

#8 Repeat steps #3 through #7 for all work groups that are to be classified.

It is important to note here that this approach assumes, at step 4, that the indexes are independent. That is, the joint (multivariate) density of all 14 indexes can be equated to

$$f(y_1, y_2, \ldots, y_{14}) = \prod_{i=1}^{14} f_i(y_i)$$

only when the indexes (y_i's) are independent. In the present case, they are not. As suggested elsewhere (Bowers, et al. 1977, pg. 57), one can consider grouping the indexes into relatively independent clusters.

The application of the Bayesian approach

- does require a fairly large developmental sample,
- does require prior "correct" classification of the developmental sample,
- does require a relatively independent set -- or collection of relatively independent subsets -- of predictors,
- does require extensive calculations, thus computer resources, for both developmental and application to new work groups.

The results of the application of this Bayesian approach are presented in the results section.

BASIC PREPARATION OF THE DATA

The data used in this study represent organizational measures on both military and civilian units. The measures of behavior within organizational life are comparable across both settings. The purpose of this section is to describe the data set used for this study.

The existing national (civilian) normative file of the Survey of Organizations (SOO) contains 5,994 groups. It represents the total body of data collected since 1966 from some 36,607 persons and 137 organizations in a broad segment of the civilian industrial population. As such, it represents many different industries, functions, and hierarchical levels.

Available also are data from two independent military samples. The first of these contains more the 787 usable groups of Navymen from whom questionnaire data on SOO indexes were collected in late 1972 and early 1973. The second contains 668 groups of Army soldiers from whom data were collected in late 1974 and early 1975. In each of these instances, in order to satisfy the need for intact units, it was decided to collect data from all members of a selected number of organizational subunits or "modules." These modules consisted of a pyramid of work groups three echelons, or tiers, tall. Thus data were collected from all members of the three organizational levels immediately below a designated "module head." Modules themselves were selected by what amounts to a stratified random sampling procedure. Methods are spelled out in greater detail in two technical reports (Michaelsen, 1973; Spencer, 1975).

Taken together, these various data sets comprise a sample of groups, to be employed in the main analyses. From the onset of the project to the present time, these various data sets have been reformatted so that all share a common format. All data have been entered into a single large file.

Measures Used

The Survey of Organizations contains in its 1974 edition 16 standard indexes. Two of these, because they have not been universally used since the start of the data bank, will be dropped. The 14 which remain will form the survey index measures to be used in the present study:

Organizational Climate

Decision Making Practices -- the manner in which decisions are made in the system: whether they are made effectively, made at the right level, and based upon all of the available information (4 item index).

Communication Flow -- the extent to which information flows freely in all directions (upward, downward, and laterally) through the organization (3 item index).

Motivational Conditions -- the extent to which conditions (people, policies, and procedures) in the organization encourage or discourage effective work (3 item index).

Human Resources Primacy -- the extent to which the climate, as reflected in the organization's practices, is one which asserts that people are among the organization's most important assets (3 item index).

Lower Level Influence -- the extent to which non-supervisory personnel and first-line supervisors influence the course of events in their work areas (2 item index).

Supervisory Leadership

Supervisory Support -- the behavior of a supervisor toward a subordinate which serves to increase the subordinate's feeling of personal worth (3 item index)

Supervisory Team Building -- behavior which encourages subordinates to develop mutually satisfying interpersonal relationships (2 item index).

Supervisory Goal Emphasis -- behavior which generates enthusiasm (not pressure) for achieving excellent performance levels (2 item index)

Supervisory Work Facilitation -- behavior on the part of supervisors which removes obstacles which hinder successful task completion, or positively, which provides the means necessary for successful performance (3 item index)

Peer Leadership

Peer Support -- behavior of subordinates, directed toward one another, which enhances each member's feeling of personal worth (3 item index).

Peer Team Building -- behavior of subordinates toward one another which encourages the development of close, cooperative working relationships (3 item index).

Peer Goal Emphasis -- behavior on the part of subordinates which stimulates enthusiasm for doing a good job (2 item index).

Peer Work Facilitation -- behavior which removes roadblocks to doing a good job (3 item index).

Satisfaction -- a measure of general satisfaction made up of items tapping satisfaction with pay, with the supervisor, with co-workers (peers), with the organization, with advancement opportunities, and with the job itself (7 item index).

The typology of work groups to be used in this study is reported in Bowers and Hausser (1977), contains 17 types, and is based on the indexes of the Survey of Organizations. The resulting types have different profiles across the indexes, with the patterns of these profiles being quite distinct.

The data set contains work groups at a variety of organizational levels. For purposes of analysis, the SOO data has been coded as representing one of the following five distinct and exhaustive classifications:

Level 4 -- Top Management. Responses of employees who report to the upper-most level of management. Includes responses from individuals in vice-presidential, major departmental heads, or equivalent positions who report to the head of the organization.

Level 3 -- Middle Management. Responses from subordinates reporting to the top management positions listed above. Includes responses of assistant department heads or general superintendents (those who head up sub-departments within major departments), technical/professional employees (with at least two subordinate levels below them), assistant controllers, etc.

Level 2 -- Second-line Supervision. Responses of foremen or equivalent positions about their supervisors (e.g., general foremen). Personnel responding here may include any supervisory personnel from the lowest (or two lowest, depending on size of organization) levels.

Level 1 -- 1st line Supervision (blue-collar). Responses of non-supervisory employees about their supervisors (e.g., foremen). Includes hourly or equivalent line workers.

Level 0 -- 1st line Supervision (white-collar). Responses of non-supervisory employees about their supervisors (e.g., clerical supervisors in accounting, personnel, etc.). May also include non-supervisory professional or technical employees.

An analysis of the present data to reflect the relative distribution of the various levels is contained in Table 1.

Another factor which distinguishes among work groups is work group size. The frequency distribution of work group size is contained in Table 2. The existence of the very small (N=1) and large work groups are

TABLE 1

DISTRIBUTION OF WORK GROUPS BY LEVEL

	Level 0	1	2	3	4	Number of Groups Classified
All work groups	35.0%	22.6%	27.1%	12.5%	2.8%	
	1,788	1,157	1,385	640	141	5,111*
Work Groups Size≤2≤N 40	31.9%	23.1%	28.3%	13.7%	3.1%	
	1,413	1,021	1,250	604	135	4,423**

*2,338 work groups have no level coding

**Of the 6,032 work groups with size between 2 and 40, 1,609 have no level coding.

TABLE 2

DISTRIBUTION OF WORK GROUPS

BY SIZE

FOR ALL WORK GROUPS

MIDPOINT	HIST%	COUNT FOR 6.GRPN	(EACH X= 14)
1.0000	18.7	1395	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
2.0000	13.3	994	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
3.0000	12.7	949	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
4.0000	10.4	776	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
5.0000	9.2	682	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
6.0000	7.5	562	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
7.0000	5.5	409	+XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
8.0000	4.6	345	+XXXXXXXXXXXXXXXXXXXXXXXXX
9.0000	3.6	269	+XXXXXXXXXXXXXXXXXXXX
10.000	2.7	203	+XXXXXXXXXXXXXXX
11.000	2.1	156	+XXXXXXXXXXXX
12.000	1.9	141	+XXXXXXXXXXX
13.000	1.5	111	+XXXXXXXX
14.000	1.0	71	+XXXXXX
15.000	.9	67	+XXXXX
16.000	.6	42	+XXX
17.000	.5	39	+XXX
18.000	.5	36	+XXX
19.000	.3	25	+XX
20.000	.4	28	+XX
21.000	.2	18	+XX
22.000	.2	14	+X
23.000	.2	16	+XX
24.000	.1	10	+X
25.000	.1	11	+X
26.000	.1	8	+X
27.000	.1	8	+X
28.000	.1	5	+X
29.000	.1	7	+X
30.000	.1	7	+X
31.000	.0	2	+X
32.000	.0	3	+X
33.000	.0	2	+X
. . .			
35.000	.0	3	+X
36.000	.0	2	+X
37.000	.0	3	+X
38.000	.1	6	+X
39.000	.0	1	+X
40.000	.0	1	+X
. . .			
45.000	.0	1	+X
. . .			
47.000	.0	1	+X
. . .			
50.000	.0	1	+X
51.000	.0	1	+X
52.000	.0	1	+X
. . .			
55.000	.0	2	+X
56.000	.0	1	+X
. . .			
58.000	.0	1	+X
59.000	.0	1	+X
. . .			
62.000	.0	1	+X
. . .			
64.000	.0	1	+X
. . .			
67.000	.0	2	+X
68.000	.0	1	+X
69.000	.0	1	+X
. . .			
80.000	.0	1	+X
. . .			
88.000	.0	1	+X
. . .			
98.000	.0	1	+X
99.000	.0	1	+X
. . .			
213.00	.0	1	+X
. . .			
489.00	.0	1	+X
TOTAL		7449	(INTERVAL WIDTH= 1.0000)

evident. The existence of work groups with N=1 most likely reflects situations where (a) other members of a small work group were absent, (b) there occurs over-specifying supervisory reporting relationships beyond those that really exist, (c) specialized survey administrations (e.g., to salaried employees only), (d) the use of samples of employees, or (e) miscoding/non-codings of work group identification numbers. Large work groups may result from certain manufacturing settings where a supervisor is directly responsible for many production people, or where, for whatever reason, middle levels of supervision were not identified.

As a partial response to the existence of very large or small work groups in the data set, some analyses are done for only those work groups with size between 2 and 40, inclusive. As shown in Table 3, there are 6,032 work groups whose size is between 2 and 40. The existence of samples of employees were especially significant in some of the military applications of the survey administration. In fact, 32% of Army work groups and 38.4% of the Navy's were of size one, where 14.7% of the civilian sample had work group N equal to one. The resulting 6,032 work groups with size 2 through 40 are thus made up of 5,095 civilian, 452 Army and 485 Navy units.

In summary, there are a total of 7,449 work groups, of which 6,944 have complete information for purposes of classification. There are 6,032 work groups with size two through 40. Of these, 5,651 have complete information. For purposes of the present analysis, the results presented will be for the 6,944 work groups and/or the 5,651 work groups.

TABLE 3

DISTRIBUTION OF WORK GROUPS BY SIZE

FOR WORK GROUP N 2≤N≤40

```
MIDPOINT    HIST%   COUNT FOR 6.GRPN   (EACH X= 10)

 2.0000     16.5     994 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 3.0000     15.7     949 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 4.0000     12.9     776 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 5.0000     11.3     682 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 6.0000      9.3     562 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 7.0000      6.8     409 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 8.0000      5.7     345 +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
 9.0000      4.5     269 +XXXXXXXXXXXXXXXXXXXXXXXXXXX
10.000       3.4     203 +XXXXXXXXXXXXXXXXXXXXX
11.000       2.6     156 +XXXXXXXXXXXXXXXX
12.000       2.3     141 +XXXXXXXXXXXXXXX
13.000       1.8     111 +XXXXXXXXXXXX
14.000       1.2      71 +XXXXXXXX
15.000       1.1      67 +XXXXXXX
16.000        .7      42 +XXXXX
17.000        .6      39 +XXXX
18.000        .6      36 +XXXX
19.000        .4      25 +XXX
20.000        .5      28 +XXX
21.000        .3      18 +XX
22.000        .2      14 +XX
23.000        .3      16 +XX
24.000        .2      10 +X
25.000        .2      11 +XX
26.000        .1       8 +X
27.000        .1       8 +X
28.000        .1       5 +X
29.000        .1       7 +X
30.000        .1       7 +X
31.000        .0       2 +X
32.000        .0       3 +X
33.000        .0       2 +X
34.000       0.        0 +
35.000        .0       3 +X
36.000        .0       2 +X
37.000        .0       3 +X
38.000        .1       6 +X
39.000        .0       1 +X
40.000        .0       1 +X

TOTAL               6032   (INTERVAL WIDTH= 1.0000)
```

RESULTS OF ANALYSES

Previous sections of this report have described the background and history of the diagnosis/classification problem, as well as approaches toward solution. In this section, numerical results of the various approaches under investigation are presented, as well as some discussion of their relative merits. The basic results of the three approaches still under consideration are presented first individually, and then a relative comparison between them.

Distance Function

The first step was to assign all work groups in the data set to one of the 17 types of the typology according to which type the work was closest to in the Euclidean sense. In order to do this, it was necessary for all 14 indexes be available for a work group. In some instances, due to the form of the SOO used and/or, most frequently, some items not being answered, all indexes were not always available. Table 4 presents the results of classifying all work groups with complete data into the 17 types of the typology. Table 4 also contains the same information for those work groups for which the work group size is between two and 40. The differences between the pattern for all work groups and those in the two to 40 range are also presented. The largest difference is in Type 1, which is the highest overall profile of scores. This difference might be expected; i.e., for those real work groups with only one subordinate, the relationship between subordinate

TABLE 4

PERCENTAGE DISTRIBUTION OF 17 TYPES

(Classified by Distance Function)

	Type																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Total Sample N=6941	4.60	5.43	12.07	8.47	7.55	6.04	9.09	5.96	3.89	4.54	7.68	3.49	7.49	4.52	3.17	3.03	2.98
Work Groups Size 2-40 N=5651	2.99	4.72	13.25	9.54	8.65	6.44	10.23	6.19	3.47	4.19	7.66	3.66	6.42	4.74	3.03	2.09	2.71
Difference in Percentages	1.61	.71	1.18	1.07	1.10	.40	1.14	.23	.42	.35	.02	.17	1.07	.22	.14	.94	.28

and manager would tend to be stronger, have greater information flow and participation in decision making, etc. The actual distribution of 1,290 classifiable work groups of size N=1 or N>40 is given in Table 5.

Two other distributions are of interest. The first is the distribution of types by level of the work group within the organization. The results of this analysis are presented in Table 6. The general pattern for this table is a higher frequency in the low numbered types for higher level management groups. This, too, is as expected. The higher score patterns (types) among higher level managerial groups have long been known to users of the SOO, and are acknowledged by the use of different norm sets (conversions from raw scores to percentiles) for each level.

The second analysis of interest is a comparison of types by the civilian/military breakdown. This analysis is reported in Table 7. The notable feature here, other than the clear difference between military and civilian patterns, is the similarity between the Navy and Army patterns. There are more high, straight line profiles in the civilian data, more low straight line and supervisor-divergent profiles in the military sample (see Bowers, 1975; Bowers & Hausser, 1977).

Decision Tree

The Decision Tree algorithm used in the present research was described in a previous section of this report. The actual implementation issues and results are reported herein. The two major implementation issues are described first.

When computing the fit of a new work group to the 17 types, the possible scores fall in the 15 point range from zero (no index score falls in any of the corresponding intervals) to 14 (all 14 index scores fall in their

TABLE 5

PERCENTAGE DISTRIBUTION OF 17 TYPES

(Classified by Distance Function)

	Type																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Total Sample N=6941	4.60	5.43	12.07	8.47	7.55	6.04	9.09	5.96	3.89	4.54	7.68	3.49	7.49	4.52	3.17	3.03	2.98
Work Groups Size 2-40 N=5651	2.99	4.72	13.25	9.54	8.65	6.44	10.23	6.19	3.47	4.19	7.66	3.66	6.42	4.74	3.03	2.09	2.71
Work Groups Size 1 or >40 N=1290	11.62	8.53	6.90	3.80	2.71	4.26	4.11	4.96	5.74	6.05	7.75	2.71	12.17	3.57	3.80	7.13	4.19

TABLE 6

DISTRIBUTION OF WORK GROUP TYPES BY LEVEL OF SUPERVISOR

		Type																
LEVEL	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1st Line White Collar 0	1672	5.7	6.3	13.6	8.6	6.8	6.5	6.5	4.2	4.7	5.6	7.1	3.9	8.5	5.2	2.4	2.5	1.9
1st Line Blue Collar 1	980	2.1	3.5	7.4	5.6	12.1	5.1	14.9	8.8	1.5	5.2	6.8	3.7	7.2	4.7	7.6	2.0	1.6
2nd Line Supervisor 2	1302	4.3	5.7	16.3	11.5	8.4	6.1	11.1	4.5	3.2	3.7	7.9	2.8	6.5	4.6	1.7	1.2	0.5
Middle Management 3	588	4.4	7.5	18.5	18.0	7.0	7.1	6.3	2.4	7.7	3.2	4.4	3.2	2.6	4.9	1.2	1.0	0.5
Top Management 4	134	9.7	11.9	20.9	17.2	6.7	1.5	1.5	1.5	13.4	3.7	3.0	1.5	1.5	6.0	0	0	0

TABLE 7

DISTRIBUTION OF WORK GROUP TYPES BY CIVILIAN OR MILITARY SETTINGS*

		Percent in Type																
	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Civilian	4725	3.1	4.9	14.6	10.6	9.4	7.0	9.9	5.6	3.8	4.3	6.6	3.7	5.7	5.4	2.8	1.6	1.0
Army	452	3.8	4.9	6.0	2.9	4.4	2.0	9.5	11.9	1.8	4.9	13.3	2.7	12.4	0.9	4.0	3.1	11.7
Navy	474	1.5	3.0	6.8	4.9	5.7	4.9	14.6	6.8	1.7	2.7	12.7	4.4	8.4	1.5	4.0	5.7	11.0

*Only for work groups with size between two and 40.

corresponding intervals). The difficulty with this fact is the possibility of ties, in particular ties for the best fit. If, for example, the 17 fit scores for a particular work group happen to be (7, 3, 6, 8, 11, 9, 6, 8, 6, 9, 10, 11, 6, 5, 3, 7, 9), respectively, then there is a tie between types 5 and 12. Conceptually, it would be possible for all types to tie. In actuality, for the intervals investigated, the ties are most frequently among two to four types. The classiciation problem is, "How do you assign work groups, whose best fit according to the decision tree algorithm, is tied among two or more types?"

Two approaches have been taken with respect to the results presented below. One approach is not to classify a group which has its best fit tied among two or more types. A second approach, knowing the correct classification, is to consider the classification to be correct if, for those work groups with best fit ties among two or more types, the correct type is among those types tied for best fit.

A second implementation issue is the selection of the intervals used in the algorithm. In the present work, while it had been decided to go with symmetric intervals about the ideal value (mean of groups for a particular type used in developing the typology), the optimum width of such intervals was unknown. The choice was to inspect a variety of intervals widths and explore their effectiveness. Tables 8, 9, 10, 11 and 12 show the frequency of correct classification and the overall percent of correct classification for the five widths, 1.4, 1.5, 1.67, 1.75 and 1.8 standard deviations either side of the ideal value. In these tables, work groups with tied best fits were counted as correctly classified if the correct type was one of those types tied for best fit.

TABLE 8

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DECISION TREE (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	295	17	1	0	0	0	0	0	0	0	0	0	5	0	0	0	1
2	74	271	9	0	0	0	0	0	4	3	0	0	16	0	0	0	0
3	17	103	467	40	5	0	0	0	83	42	5	0	76	0	0	0	0
4	0	1	25	340	40	14	0	0	59	39	42	4	4	20	0	0	0
5	0	1	5	32	275	25	22	0	4	63	27	32	2	35	1	0	0
6	0	1	2	5	8	305	21	12	3	3	19	5	0	26	1	1	7
7	0	0	0	0	16	42	361	47	0	6	24	30	1	8	55	5	36
8	0	0	0	0	0	5	5	299	0	1	1	4	0	1	21	48	29
9	22	13	2	12	0	2	0	0	211	0	0	1	6	0	0	0	1
10	3	1	2	0	1	0	1	1	1	277	3	8	8	2	4	2	1
11	2	1	1	7	6	20	5	5	7	13	406	1	29	0	8	5	17
12	2	3	1	0	3	0	1	2	0	15	1	192	0	6	15	1	0
13	40	10	7	1	6	2	1	0	0	19	16	0	414	0	3	0	1
14	4	2	1	4	4	22	3	8	11	2	1	20	0	229	0	1	2
15	0	0	0	0	1	0	0	1	0	2	1	3	0	0	198	5	9
16	0	0	0	0	1	2	2	12	0	2	1	4	0	1	13	165	7
17	0	1	0	0	0	1	2	3	0	0	1	0	1	0	4	4	190

NUMBER CORRECT 4895 PERCENT CORRECT 70.5

TABLE 9

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DECISION TREE (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	301	16	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	82	268	7	0	0	0	0	0	4	2	0	0	14	0	0	0	0
3	19	94	482	43	3	0	0	0	83	39	5	0	70	0	0	0	0
4	0	0	35	344	31	8	1	0	56	40	45	2	3	23	0	0	0
5	0	0	4	36	270	29	22	0	3	57	43	27	2	30	1	0	0
6	0	1	2	4	6	304	15	10	2	2	30	4	0	35	0	1	3
7	0	0	0	0	11	30	385	48	0	8	34	42	0	7	40	2	24
8	0	0	0	0	0	5	5	300	0	0	1	5	0	3	21	53	21
9	29	3	3	6	1	2	1	0	210	1	7	0	6	1	0	0	0
10	4	1	4	2	0	0	1	1	1	270	9	6	9	2	4	0	1
11	3	1	3	4	2	18	3	5	4	7	444	0	18	1	5	1	14
12	2	3	1	0	2	0	0	3	0	8	1	195	1	14	11	1	0
13	48	5	7	3	1	1	0	0	0	9	22	0	422	0	1	0	1
14	3	0	1	4	2	27	6	6	7	3	3	18	0	232	0	2	0
15	0	0	0	0	1	0	0	3	1	2	1	3	0	0	192	6	11
16	0	0	0	0	1	2	1	9	0	3	1	3	0	0	9	168	13
17	0	1	0	0	0	2	1	3	0	0	4	1	0	0	3	7	185

NUMBER CORRECT 4972 PERCENT CORRECT 71.6

TABLE 10

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DECISION TREE (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	297	18	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0
2	73	273	4	0	0	0	0	0	2	5	0	0	20	0	0	0	0
3	11	90	497	39	2	0	0	0	76	46	6	1	69	1	0	0	0
4	0	0	29	350	34	14	1	0	49	40	41	4	5	21	0	0	0
5	0	0	2	23	295	21	26	0	4	63	33	25	3	29	0	0	0
6	0	1	1	4	3	300	20	8	5	7	30	2	0	32	0	2	4
7	0	0	0	0	7	36	400	53	0	10	21	34	0	5	40	5	20
8	0	0	0	0	0	5	9	287	0	1	0	6	0	1	17	64	24
9	18	5	4	8	1	0	1	0	219	0	3	0	9	2	0	0	0
10	2	2	1	1	0	0	0	1	0	285	2	2	14	1	3	1	0
11	1	1	1	6	2	14	2	4	5	10	442	1	22	3	3	5	11
12	3	1	0	0	2	1	1	0	0	9	0	202	0	12	10	1	0
13	23	4	1	0	0	0	0	0	0	21	11	0	458	0	1	0	1
14	2	1	1	3	3	21	2	3	5	3	1	17	0	250	1	1	0
15	0	0	0	0	0	0	2	2	0	2	0	3	2	0	194	11	4
16	0	0	0	0	0	1	0	7	0	1	3	5	0	0	4	187	2
17	0	0	0	0	0	1	0	7	0	1	3	2	2	0	5	7	179

NUMBER CORRECT 5115 PERCENT CORRECT 73.7

TABLE 11

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DECISION TREE (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	304	13	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
2	74	268	6	0	0	0	0	0	1	7	0	0	21	0	0	0	0
3	13	67	529	39	0	0	0	0	75	43	8	1	60	3	0	0	0
4	0	0	22	371	28	13	1	0	45	38	42	3	4	21	0	0	0
5	0	0	2	22	304	20	21	0	1	71	32	21	1	28	1	0	0
6	0	2	0	5	2	311	10	9	4	7	26	2	0	37	0	1	3
7	0	0	0	0	11	36	418	41	0	11	20	32	0	7	28	3	24
8	0	0	0	0	0	8	8	293	0	1	0	4	1	0	14	58	27
9	16	4	2	7	1	0	1	0	223	0	1	0	11	4	0	0	0
10	2	2	2	1	0	0	1	1	0	282	1	5	10	1	5	2	0
11	0	1	0	4	1	14	2	3	5	10	446	1	22	4	3	3	14
12	3	1	0	0	2	0	2	0	0	11	0	198	0	14	9	2	0
13	19	2	2	1	0	0	0	0	0	18	11	0	464	0	2	0	1
14	2	1	0	1	1	24	1	3	3	4	2	20	0	251	1	0	0
15	0	0	0	0	0	0	4	1	0	1	1	2	2	0	194	8	7
16	0	0	0	0	0	2	1	4	0	1	2	5	0	0	2	182	11
17	0	0	0	0	0	0	0	7	0	0	1	1	2	0	3	5	188

NUMBER CORRECT 5226 PERCENT CORRECT 75.3

TABLE 12

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DECISION TREE (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	302	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	71	273	9	0	0	0	0	0	1	6	0	0	17	0	0	0	0
3	12	68	531	39	2	0	0	0	76	44	8	1	56	1	0	0	0
4	0	0	26	384	22	13	2	0	35	39	38	2	4	23	0	0	0
5	0	0	2	18	319	28	23	0	1	60	27	21	2	22	1	0	0
6	0	2	1	2	2	315	14	11	3	4	24	2	0	33	0	1	5
7	0	0	0	0	7	34	430	41	0	12	18	30	0	6	26	2	25
8	0	0	0	0	0	6	6	303	0	1	0	6	1	1	11	54	25
9	16	5	2	7	1	0	0	0	222	0	1	0	10	5	0	0	1
10	2	2	3	1	0	0	3	1	0	278	3	5	9	1	5	2	0
11	2	1	0	2	5	19	2	2	5	9	435	1	25	3	3	3	16
12	2	1	0	0	2	0	5	0	0	8	0	201	0	14	8	1	0
13	22	2	1	1	0	0	0	0	2	18	11	0	460	0	3	0	0
14	2	1	0	1	2	22	0	3	4	4	2	20	0	251	1	1	0
15	0	0	0	0	0	0	5	1	0	0	1	3	2	0	195	7	6
16	0	0	0	0	0	2	1	6	0	1	2	4	0	0	2	180	12
17	0	0	0	0	0	0	0	8	0	0	1	0	2	0	4	4	188

NUMBER CORRECT 5267 PERCENT CORRECT 75.9

Inspection of Tables 8 through 12 show an increasing percentage accuracy with wider intervals. This pattern is somewhat misleading in that there is a corresponding, non-linear (greater than linear) increase in the frequency of work groups with tied best fit scores. The summary of these results across the five tables can be seen in Table 13.

In selecting a standard interval width, the primary concern would seem to be one of ties. As shown in Table 13, the percent accuracy of classification does not effectively change if one classifies only those work groups for which there are not ties for best type. This result needs to be interpreted, however, in light of the realization that there are fewer work groups to classify with the larger intervals than the smaller. Thus, for intervals of ± 1.4 standard deviations, there were (6,941-2,190 with ties =) 4,751 work groups to classify, whereas with intervals of ± 1.8 standard deviations, there were (6,941-3,001 with ties =) 3,940 work groups to classify. Thus, the larger interval classified (57.5% of 3,940 =) 2,266 correctly, whereas the smaller classified (56.9% of 4,751 =) 2,705 correctly. These results indicate selecting one of the smaller intervals.

A second concern, however minor, is the distribution of the number of ties. In practice, if approximately 30 to 35% of work groups have tied best fit scores, the classification problem is different if most work groups have only two types which tie or, say, four or five types are tied. In the former, for treatment purposes, one might explore the advantage of both suggested treatments, whereas in the latter, one would still be selecting among four to five treatments. Table 14 demonstrates that the number of tied types is principally two, especially in the shorter intervals.

TABLE 13

ACCURACY OF CLASSIFICATION BY THE DECISION TREE

(All Work Groups)*

	Number of Standard Deviation Units Used On Either Side of the Mean				
	1.40	1.50	1.67	1.75	1.80
Percent Accurate Classification	70.5	71.6	73.7	75.3	75.9
Percent of work groups which had best fit in two or more work groups	31.6	33.2	38.1	41.7	43.2
Percent of work groups without ties that were accurately classified	56.9	57.6	57.5	57.7	57.5

*Based on 6941 work groups, regardless of work group size.

TABLE 14

DISTRIBUTION OF NUMBER OF TIED TYPES

Interval About Mean	Number of Types That Tied							
	2	3	4	5	6	7	8	N
$\pm$ 1.40σ	70.8	21.1	6.2	1.1	0.5	0.1	0.0	2190
$\pm$ 1.50σ	69.2	22.6	5.6	1.7	0.7	0.1	0.0	2302
$\pm$ 1.67σ	67.7	23.1	7.1	1.2	0.6	0.1	0.0	2646
$\pm$ 1.75σ	65.0	23.7	8.6	1.7	0.7	0.2	0.1	2891
$\pm$ 1.80σ	62.3	24.9	9.5	2.3	0.7	0.3	0.1	3001

Another feature of concern is whether the ties are distributed proportionately across actual types or not. Table 15 presents the distribution pattern of the work groups with tied types for best fit. By comparison with the total distribution, one sees that there are somewhat more ties among the relatively frequent correct types and fewer ties among the more infrequent correct types. The diagnostic implications are that the decision tree process will do somewhat better in identifying the rarer types than the more common types.

A fourth feature is the degree of compatability between a particular work group's index values and the intervals, which has been called the fit. The analysis of best fit values, for only those work groups that had tied types, is presented in Table 16. It shows that the average fit becomes higher with larger intervals, as expected. For differentiation purposes, it also suggests the use of smaller intervals.

As a consequence of the results presented above, it was decided to select the intervals given by ± 1.5 standard deviations. For this size, the accuracy of classification was 71.6% with 33.2% of the work groups having tied best fit types. The accuracy was 57.6% when classifying only those work groups that did not have tied best fit types. In future references to the decision tree algorithm, the interval used will be ± 1.5 standard deviations.

A last feature of the classification accuracy of this procedure is the accuracy with which it classifies within each type. That is, given that a work group is of a certain actual type, is the probability of being correctly classified the same as it is if it were a different actual type? The results, computed from the figures in Table 9, are presented in Table 17. As can be seen, the percentages of correct classification are not uniform across types. The greatest difficulty is for types 3, 4, and 5. These types represent

TABLE 15

PERCENTAGE DISTRIBUTION OF WORK GROUPS WITH TIED BEST FIT BY CORRECT TYPE

Interval About Mean	Correct Type 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$\pm$ 1.40σ	1.2	3.9	15.0	11.6	10.9	7.5	11.1	6.5	3.2	3.4	6.2	3.3	5.3	4.5	1.8	2.7	1.9
$\pm$ 1.50σ	1.2	3.6	15.9	11.3	10.9	6.8	11.9	6.8	3.2	3.4	5.7	3.1	5.6	4.4	1.8	2.5	2.0
$\pm$ 1.67σ	1.5	3.7	16.3	11.5	11.3	7.0	12.0	6.4	2.5	3.6	5.7	3.3	4.7	4.2	2.0	2.4	1.9
$\pm$ 1.75σ	1.7	3.9	16.2	11.3	11.1	7.2	12.5	6.6	3.3	3.1	6.0	2.8	4.1	4.0	2.0	2.5	1.9
$\pm$ 1.80σ	1.6	4.1	15.7	11.9	11.5	7.0	12.2	7.0	3.1	3.2	5.7	2.9	4.2	3.8	1.9	2.3	1.9
Distribution of Work Groups	4.6	5.4	12.1	8.5	7.6	6.0	9.1	6.0	3.9	4.5	7.7	3.5	7.5	4.5	3.2	3.0	3.0

TABLE 16

AVERAGE BEST FIT*

Interval About Mean	Mean Fit **	Percent of Groups With Maximum Fit**
$\pm$ 1.40σ	11.104	7.1
$\pm$ 1.50σ	11.603	14.2
$\pm$ 1.67σ	12.411	32.4
$\pm$ 1.75σ	12.658	40.9
$\pm$ 1.80σ	12.827	46.9

**The fit is the number of indexes for which the work group data falls within the designated range; the maximum is 14.

**These figures are for those work groups which had best fit in two or more types (i.e., tied for which type it is to be classified).

TABLE 17

DECISION TREE ACCURACY BY TYPE

Correct Type	Percent Correct
1	94.4
2	71.1
3	57.5
4	58.5
5	51.5
6	72.6
7	61.0
8	72.5
9	77.8
10	85.7
11	83.3
12	80.6
13	81.2
14	73.9
15	87.3
16	80.0
17	89.4
Overall	71.6

SOO score profiles which fall close to the 50th percentile. In examining Table 9, one sees that the predominant pattern is to assign these straight line types into the non-straight line types (the latter are types 9 to 17). Most of the divergent profiles begin with score patterns in this 45th to 65th percentile range, (the range covered by types 3, 4, and 5) which may explain the difficulty of correctly classifying these straight line types, while not having nearly the difficulty with the other straight line types which are higher or lower.

Discriminant Function

The discriminant function analysis reported here involves classifying work groups on the basis of three super indexes: organizational climate, supervisory leadership, and peer leadership. The reasons behind the consolidation of the regular SOO indexes into these three super indexes were given in an earlier section. The total data set was divided into two random subsets, each consisting of 3,016 work groups whose size ranged from two to 40. The discriminant functions were generated from one sample and then applied to the other sample.

The pattern of classification of this procedure is given in Table 18. The percentage of correct classification is based on 2,823 work groups, (2,828 = 3,016-188 work groups without "correct" classification).

The discriminant functions were also applied to the developmental sample (i.e., the data set from which they were generated.) The results are given in Table 19. The remarkable consistency of the accuracy of prediction (77.0% & 77.4%) indicate substantial stability across data sets. This is significant and indicates portability of such equations as well as decreasing the importance of a large data set for developmental purposes.

TABLE 18

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DISCRIMINANT FUNCTION (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	69	15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	12	112	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	37	238	27	0	0	0	0	50	3	0	0	4	0	0	0	0
4	0	0	4	217	20	5	0	0	16	3	3	0	1	5	0	0	0
5	0	0	0	23	165	11	5	0	0	14	3	7	0	11	0	0	0
6	0	0	0	2	18	145	12	1	1	0	1	0	0	0	0	0	0
7	0	0	0	0	6	17	236	12	0	0	4	18	0	1	3	0	9
8	0	0	0	0	0	0	3	127	0	0	0	0	0	0	2	19	12
9	0	2	18	5	0	0	0	0	75	0	0	0	0	0	0	0	0
10	0	0	2	1	4	0	1	0	0	100	1	11	2	0	0	0	0
11	0	0	0	22	7	23	2	0	3	1	183	0	7	0	0	0	3
12	0	0	0	0	3	0	0	1	0	2	0	98	0	1	0	0	0
13	0	3	9	4	0	0	0	0	0	12	6	0	124	0	0	0	0
14	0	0	0	1	9	11	3	0	2	0	0	5	0	97	0	1	0
15	0	0	0	0	0	0	1	0	0	0	0	3	0	0	66	3	7
16	0	0	0	0	0	0	0	8	0	0	0	0	0	0	2	46	0
17	0	0	0	0	0	2	4	7	0	0	3	0	0	0	0	2	79

NUMBER CORRECT 2177. PERCENT CORRECT 77.0

TABLE 19

DEVELOPMENTAL SAMPLE
THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DISCRIMINANT FUNCTION (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	75	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	7	132	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0
3	0	35	249	32	0	0	0	0	61	11	0	0	2	0	0	0	0
4	0	0	3	207	22	3	0	0	15	8	1	0	0	6	0	0	0
5	0	0	0	25	174	6	2	0	0	10	5	8	0	20	0	0	0
6	0	0	0	3	16	151	10	1	0	0	1	0	0	2	0	0	0
7	0	0	0	0	2	19	210	8	0	0	2	15	0	3	4	0	9
8	0	0	0	0	0	4	4	146	0	0	0	1	0	0	1	25	6
9	0	1	15	8	0	1	0	0	71	0	0	0	0	0	0	0	0
10	0	0	3	1	1	0	0	0	0	98	1	4	7	0	0	0	0
11	0	0	1	10	5	8	0	0	3	2	149	0	3	0	0	0	1
12	0	0	0	0	2	0	0	0	0	0	0	90	0	9	1	0	0
13	3	7	12	2	0	0	0	0	1	7	19	0	154	0	0	0	0
14	0	0	0	5	10	5	1	1	3	0	0	4	0	110	0	0	0
15	0	0	0	0	0	0	2	0	0	0	0	0	0	0	78	8	3
16	0	0	0	0	0	0	0	14	0	0	0	0	0	0	0	48	0
17	0	0	0	0	0	0	6	6	0	0	1	0	0	0	0	0	43

NUMBER CORRECT 2185. PERCENT CORRECT 77.4

The combination of the developmental and test samples are reported in Table 20. For Table 21, the total data set, including those work groups of size one or greater than 40 which were previously omitted, was submitted for discriminant function classification. The decline in accuracy (albeit, only from 77.2% to 76.6%) indicates that the odd sized work groups (mostly N=1) are not classified as accurately as those with size between two and 40.

The last analysis of these figures was to examine the accuracy of correct classification by "correct" type. The results of these calculations are given in Table 22. The two types with the lowest accuracy figures are types three and five. In both types, the adjacent straight line profiles, combined, would appear to be nearly as attractive alternatives as the non-straight line profiles (types nine to 17).

Bayesian Method

The Bayesian approach, as discussed earlier, involved several steps (see page 46). Important to the method is the concept of independence. This independence is not with respect to work groups, but with respect to the SOO indexes which are the "predictor variables." This section discusses the results in applying the Bayesian method to the organizational diagnosis problem.

As discussed by Bowers, et al. (1977), there are two traditional approaches to handling the independence issue: (1) ignoring the lack of independence, and (2) clustering the predictor variables into relatively independent subsets. The first approach raised significant computational problems. To describe them, it is helpful to go through the implementation activities.

TABLE 20

COMBINATION: TEST AND DEVELOPMENT
THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DISCRIMINANT FUNCTION (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	144	24	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	19	244	3	0	0	0	0	0	0	0	0	0	1	0	0	0	0
3	0	72	487	59	0	0	0	0	111	14	0	0	6	0	0	0	0
4	0	0	7	424	42	8	0	0	31	11	4	0	1	11	0	0	0
5	0	0	0	48	339	17	7	0	0	24	8	15	0	31	0	0	0
6	0	0	0	5	34	296	22	2	1	0	2	0	0	2	0	0	0
7	0	0	0	0	8	36	446	20	0	0	6	33	0	4	7	0	18
8	0	0	0	0	0	4	7	273	0	0	0	1	0	0	3	44	18
9	0	3	33	13	0	1	0	0	146	0	0	0	0	0	0	0	0
10	0	0	5	2	5	0	1	0	0	198	2	15	9	0	0	0	0
11	0	0	1	32	12	31	2	0	6	3	332	0	10	0	0	0	4
12	0	0	0	0	5	0	0	1	0	2	0	188	0	10	1	0	0
13	3	10	21	6	0	0	0	0	1	19	25	0	278	0	0	0	0
14	0	0	0	6	19	16	4	1	5	0	0	9	0	207	0	1	0
15	0	0	0	0	0	0	3	0	0	0	0	3	0	0	144	11	10
16	0	0	0	0	0	0	0	22	0	0	0	0	0	0	2	94	0
17	0	0	0	0	0	2	10	13	0	0	4	0	0	0	0	2	122

NUMBER CORRECT 4362. PERCENT CORRECT 77.2

TABLE 21

ALL WORK GROUPS
THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DISCRIMINANT FUNCTION (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	288	29	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
2	48	321	3	0	0	0	0	0	0	1	0	0	4	0	0	0	0
3	0	80	529	66	0	0	0	0	139	17	0	0	7	0	0	0	0
4	0	0	7	451	50	10	0	0	36	11	7	0	1	15	0	0	0
5	0	0	0	56	360	17	8	0	0	28	9	15	0	31	0	0	0
6	0	0	0	5	34	342	26	3	3	0	4	0	0	2	0	0	0
7	0	0	0	0	9	44	477	22	0	0	6	36	0	4	7	0	26
8	0	0	0	0	0	6	8	314	0	0	0	1	0	0	3	59	23
9	3	3	40	17	0	2	0	0	202	0	0	0	3	0	0	0	0
10	0	2	5	2	10	0	1	0	0	256	4	20	12	0	3	0	0
11	0	0	1	39	14	33	3	0	13	3	411	0	12	0	0	0	4
12	0	0	0	0	7	0	0	2	0	2	0	214	0	13	4	0	0
13	9	18	26	6	0	0	0	0	2	25	34	0	400	0	0	0	0
14	0	0	0	8	21	22	4	3	5	1	0	11	0	238	0	1	0
15	0	0	0	0	0	0	3	0	0	0	0	3	0	0	189	14	11
16	0	0	0	0	0	0	0	34	0	0	0	1	0	1	4	170	0
17	0	0	0	0	0	3	14	19	0	0	9	0	0	0	0	5	157

NUMBER CORRECT 5319. PERCENT CORRECT 76.6

The computer program that operationalizes the steps given earlier (page 46) required the estimates of the two parameters, α and β, for the beta distributions which were used to model the actual (but unknown) distributions of each index within each type. The α values ranged from just over one to nearly 28, while β values were from near zero to 20.5. The resulting constant terms in the densities (the ratio of gamma terms in equation (1) on page 45) ranged from two to two x 10^{31}. The result of having numbers of this magnitude, when computing the odds-likelihood ratio (given by equation 4, step 4, page 46) was obtaining values which exceeded the normal storage capabilities of computers (10^{72} or 10^{-72}). Thus, the algorithm was not possible to implement in its present form. This is primarily due to the lack of independence between indexes, and partially due to the number of predictor variables.

In the investigation of possible clusters of indexes which have a reasonable conceptual basis, the pairwise correlations remained above 0.5. The multiple relationships would be even larger. The inability to meet the independence criterion of this technique remains a major obstacle. It also seems highly unlikely that any set of predictors (not just SOO measures) chosen to reflect organizational functioning will meet the independence criterion.

While multivariate models not requiring independence are available for use as prior distributions, the sample N necessary to estimate all the parameters would exceed greatly the relatively large N=7,000 that is presently available.

It seems unlikely, therefore, that implementation of Bayesian classification procedures, within the constraints of present application techniques,

will prove viable to data-based organizational diagnosis. It would appear that it rests with the mathematical statisticians to develop (and have distributed) techniques for using Bayesian approaches to highly interdependent, multi-faceted data.

TABLE 22

DISCRIMINANT FUNCTION ACCURACY BY TYPE

Type	Percent Accurately Classified		
	Total Sample	Test Sample	Developmental Sample
1	90*	81	89
2	85	90	92
3	63	66	64
4	77	79	76
5	69	69	70
6	82	81	82
7	75	77	77
8	76	78	78
9	75	75	74
10	81	82	85
11	77	73	82
12	88	93	88
13	77	78	75
14	76	75	79
15	86	83	86
16	81	82	77
17	76	81	77

*The total sample values are not the average of the two sample values as the former only includes work groups of size one and those larger than N=40.

ADDITIONAL CRITERIA

The results presented thus far have reflected one primary concern of classification procedures, that being, proportion of correct classification. There are other accuracy oriented criteria as well as non-accuracy based criteria which are also of interest.

One of the accuracy based criteria is the degree to which the clusters of work groups (or other units of analysis) generated by the classification techniques replicate the typology. Another criteria re-defines the concept of closeness of a work group to the definitional values of the typology. Still others have examined the cost (loss) of misclassifying a particular work group into the "wrong" type. These accuracy oriented criteria are used to assess the organizational diagnosis procedures under consideration.

In an earlier report in this series (Bowers, et al, 1977), several concepts relating to criteria other than accuracy based criteria were discussed as possible alternative approaches to evaluating classification procedures. Three such non-accuracy based criteria are considered here. One is whether the amount of information used in making a classification can be reduced. Another is whether there are considerable differences in the amount of data necessary to develop the procedures. Finally, the ease in calculating these data based classifications, whether they can be done "on-site," is a major feature.

Reproduce the Typology

If one thinks of a typology which involves k variables and n classes as n points in k-dimensional space, one approach to evaluating a classification procedure is to see how well the clusters of classified units gather

about the n points. To the extent that the clusters don't group about the definitional types, one can conclude that the lack of fit is due to either the classification procedure or the inappropriateness of the typology itself (or a mixture of the two). The discussion here is primarily focused at the classification procedures, but also implications regarding the typology are presented.

Two major features of the fit of clusters to the typology are the proximity of the vector of means of a cluster to the corresponding definitional values of that type, across all types, and the degree of dispersion within each cluster. To measure the proximity, the vector of index means was computed for each of the 17 clusters which resulted from each of the three classification procedures (distance function, discriminant function, and decision tree). Then the distance between cluster mean vectors and the definitional values* of the respective types was calculated. The results are given in Table 23.

Two aspects of this analysis are relevant. First, looking only at the relative magnitude of the distances, first across classification procedures and then down the types, one sees that the distance procedure has the best pattern of smallest distances by index. The decision tree procedure has the second best pattern. The former is as expected. The latter, while possibly surprising, is due to the nature of the algorithm which assigns work groups to types by fitting them to the definitional values.

---

*The values which were used as the definitional values of the typology were results of the development of the typology (Bowers & Hausser, 1975). In particular, they were the simple means of the cluster means from each of the three samples used to develop the typology.

TABLE 23

DISTANCE BETWEEN TYPE CLUSTERS AND VALUES

GIVEN BY THE TYPOLOGY

	Classified by		
Type	Distance Function	Discriminant Function	Decision Tree
1	.3573	.4655	.6581
2	.4886	.6836	.6418
3	.2270	.2518	.2112
4	.2128	.3100	.2547
5	.1746	.2567	.1775
6	.4118	.5794	.4308
7	.2258	.2917	.2537
8	.3367	.4347	.4567
9	.3859	.7001	.5384
10	.3650	.4818	.5234
11	.4114	.5769	.4463
12	.3430	.4768	.4693
13	.4451	.5547	.4935
14	.2562	.3312	.2781
15	.3057	.3782	.5337
16	.5715	.8907	.7217
17	.4546	.6679	.6012

The interpretation of these distances is confounded by the fact that the different procedures assigned different work groups, and thus, different numbers of work groups, to the individual types (see Table 24). For example, while the decision tree procedure has the smallest mean distance to the Type 3 definitional values, it also classified 244 fewer groups as Type 3 than did the distance function (the 244 groups were thus said to be misclassified). Hence, absolute comparisons in Table 23 must be made with caution.

A second feature of the values given in Table 23 is the interpretability of the numbers. For example, how does one interpret a distance of .3573 in 14 dimensional space? As a partial answer to this dilemma, the distance between two pairs of types (using the definitional values) were computed. They were:

distance (Type one to Type six) = 1.5334,

distance (Type five to Type six) = 1.0218.

Thus, the distance of .3573 between the distance function cluster one and the defining values is, at most, 23 percent of the distance toward Type 2 from Type 1. (Given the geometry of 14 dimensional space, directionality is not linear and thus the direction of the distance is difficult to interpret.)

Again referring to Table 23, one notices considerable variation across types with respect to the distance from the definitional values. This variation, however, is not particularly greater among the divergent types (Types nine to 17) than among the straight line types (Types one to eight). Given that the dispersion, as will be shown shortly, is not particulary

TABLE 24

NUMBER OF WORK GROUPS PER TYPE*

	Classified by		
Type	Distance Function	Discriminant Function	Decision Tree
1	169	166	217
2	267	353	312
3	749	557	505
4	539	595	415
5	489	464	304
6	364	411	384
7	578	502	403
8	350	332	329
9	196	301	305
10	237	271	357
11	433	383	532
12	207	264	265
13	363	306	419
14	268	265	319
15	171	157	224
16	118	152	165
17	153	172	196

N=5,651

*For only work groups with N, $2 \leq N \leq 40$

greater for those types with larger distance values, changes in the definitional values of the typology for those types are suggested.

The second feature of fit is the degree of dispersion within the clusters. An (unreal) ideal model would have all cluster members falling exactly at the same point in space. The actual dispersion was computed for each index within each type, and averaged across indexes. These average values are shown in Table 25. Also given is the weighted average of standard deviations from the three samples that were used to develop the typology. The general pattern can be seen to be that all three classification schemes assign groups to clusters whose standard deviations are smaller than for the developmental sample of the typology (three subsamples).

Comparison of the values in Table 25 is simplified if one assigns ranks to the four values given for each type. These values are given in Table 26. The pattern of ranks for discriminant function shows that it is the least variable within clusters (recall, however, it was the furthest distance from the definitional values). However, returning to Table 25, one notices that most often the difference in average variability between clusters given by the distance function and the discriminant function is less than .01. Effectively, it can be said that the variance within clusters is approximately the same for all three techniques, all of which are better than for the developmental sample.

In summary, the distance function seems to reproduce the typology most effectively (as expected). While there is some difference between the distance of the vectors of means of the clusters and the definitional values for the three classification procedures, the variability within clusters is approximately the same across procedures.

TABLE 25

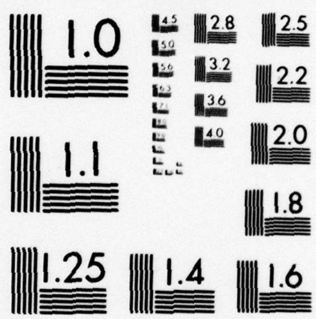
AVERAGE (ACROSS INDEXES) STANDARD DEVIATIONS BY TYPE

	Type																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Developmental Sample	.4218	.4127	.3493	.3201	.3343	.2986	.3852	.3911	.4850	.4204	.4036	.4828	.4732	.3529	.5060	.4679	.4307
Distance Function	.36936	.35877	.33039	.31579	.31782	.35813	.34641	.39715	.36513	.40901	.40046	.39157	.40860	.37579	.44491	.44367	.43761
Discriminant Function	.37104	.35188	.31538	.31903	.31400	.36202	.33922	.38995	.36923	.38898	.40722	.39414	.40478	.36813	.42723	.44532	.42269
Decision Tree	.46794	.34401	.32132	.30019	.30848	.36016	.32003	.38647	.37425	.40691	.41447	.38907	.40812	.38448	.45650	.46445	.47606

AD-A065 289 MICHIGAN UNIV ANN ARBOR INST FOR SOCIAL RESEARCH F/G 5/11
ORGANIZATIONAL DIAGNOSIS: ISSUES AND METHODS.(U)
FEB 79 A S DAVENPORT, D G BOWERS N00014-77-C-0096
UNCLASSIFIED NL

2 OF 2
AD
A065289

END
DATE
FILMED
4 -79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TABLE 26

RANKS OF DISPERSION WITHIN TYPES

Type	Typology Developmental Sample	Distance Function	Discriminant Function	Decision Tree
1	3	1	2	4
2	4	3	2	1
3	4	3	1	2
4	4	2	3	1
5	4	3	2	1
6	1	2	4	3
7	4	3	2	1
8	4	3	2	1
9	4	1	2	3
10	4	3	1	2
11	2	1	3	4
12	4	2	3	1
13	4	3	1	2
14	1	3	2	4
15	4	2	1	3
16	4	1	2	3
17	2	3	1	4

All three procedures seem to reproduce the typology with some effectiveness. One implication is that the typology is not just developmental sample based, and does have usefulness in a broader arena.

Zero-One Measure

As shown in the Introduction of this report, two differt work groups may be equally distant from a particular type, but have quite different closeness properties. A second way to measure closeness (other than the distance function) is to count the number of indexes of a particular work group which are "far away" from the type values. The arithmetic of this counting function for each work group can be represented by:

$$C = \sum_{i=1}^{14} C_i$$

where C_i is given by

$$C_i = \begin{cases} 1, & \text{if} |Y_i - X_i| \geq s, \\ 0, & \text{if} |Y_1 - X_i| < s, \end{cases}$$

for i=1,2,...,14. Here s is some defining value representing "significant difference," and $\underline{Y}$ and $\underline{X}$ are the vectors corresponding to the definitional values of a type and a work group's actual scores, respectively.

In order to compute C for each work group, average them within types, and then average them across types, s was given to be some weight times the standard deviation of the ith index (averaged across types). The results for different weights ranging from .5 to 1.25 are given in Table 27. As would be expected, the average number of differences decline as the significant difference value increases. Also, there appears to be no significant difference among the classification procedures with respect to this criterion (for the weight values given).

TABLE 27

AVERAGE NUMBER OF INDEXES PER WORK GROUP

WHICH DIFFER FROM ASSIGNED TYPE*

	Weights				
	.5	.75	.8	1.0	1.25
Distance Function	8.552	6.222	5.876	4.482	3.029
Discriminant Function	8.638	6.312	5.978	4.562	3.114
Decision Tree	8.694	6.337	5.981	4.482	2.941

*Differing on index i is said to occur if the observed value differs from the corresponding index of the defining profile by more than the <u>weight</u> times the standard deviation of that index.

5,651 work groups

There is an interesting relationship between this zero-one counting criterion and the decision tree algorithm. This zero-one method counts the number of times an index is considerably different from the defining value. The decision tree algorithm, as used herein, effectively counts the number of times an index is close to the defining value. Thus, if the weight had been 1.5, then the decision tree algorithm would have been effectively the same as classifying by minimizing the zero-one count.

The last feature to be gleaned from Table 27 is that there seems to be considerable deviation between most work group index values and the corresponding index values of the definitional types juding from the magnitude of the average values given. However, the size of most index standard deviations is slightly less than .4, so, for example, when the weight is one, 4 1/2 indexes (on the average) fall more than .4 from the corresponding definitional value.

While not shown here, there was little or no difference between types with respect to the average number of differing work group values.

Severity of Misclassification

As indicated earlier, errors of classification may vary quite widely in their impact upon the group so classified. The impact stems, of course, from the consequences of applying to the classified group a treatment appropriate to the type of misclassification. If, for example, the group should have received survey feedback but instead receives data handback, the consequences might be on absence of the positive change that would normally have occurred, or even outright deterioration. Since appropriate treatment is determined by the most favorable change consequences (as far as our data carry us), in no case would the result be more favorable than that indicated for the appropriate treatment.

By Multiple Discriminant Function

To examine this, we returned to the table of change consequences contained in the original typology analysis (Bowers & Hausser, 1977). For classification by the discriminant function, accuracy (compared to our criterion, the distance function used in the original typology) is displayed in Table 28. Accurate classification is contained in those frequencies falling in the upper left to lower right diagonal. All off-diagonal cases are therefore instances of misclassification. Of the total of 6,944 cases, 5,319 (77 percent) are classified correctly. One thousand six hundred twenty five (or 23 percent) are misclassified.

Of the 5,319 classified correctly, 5,149 are in other than Type 16, for which type no treatment evidence was available in the original analysis. To these 5,149, we applied a simple coding scheme:

4 = ++ change in the original findings
3 = + change in the original findings
2 = 0 change in the original findings
1 = - change in the original findings
0 = -- change in the original findings

The mean change for the 5,149 correctly classified cases would therefore have been 3.10 (or "+.")

Of the 1,625 misclassified cases, 1,157 were in types (correct or misclassified) for which both correct classification change codes and misclassification change codes existed. To assess the impact of misclassification, the following coding scheme was used:

4 = Assigned to treatment where result was (--) instead of correct one which would have been (++.)

3 = Assigned to treatment where result was (--) instead of correct one which would have been (+.)

or

Assigned to treatment where result was (-) instead of correct one which would have been (++.)

TABLE 28

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DISCRIMINANT FUNCTION (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	288	29	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
2	48	321	3	0	0	0	0	0	0	1	0	0	4	0	0	0	0
3	0	80	529	66	0	0	0	0	139	17	0	0	7	0	0	0	0
4	0	0	7	451	50	10	0	0	36	11	7	0	1	15	0	0	0
5	0	0	0	56	360	17	8	0	0	28	9	15	0	31	0	0	0
6	0	0	0	5	34	342	26	3	3	0	4	0	0	2	0	0	0
7	0	0	0	0	9	44	477	22	0	0	6	36	0	4	7	0	26
8	0	0	0	0	0	6	8	314	0	0	0	1	0	0	3	59	23
9	3	3	40	17	0	2	0	0	202	0	0	0	3	0	0	0	0
10	0	2	5	2	10	0	1	0	0	256	4	20	12	0	3	0	0
11	0	0	1	39	14	33	3	0	13	3	411	0	12	0	0	0	4
12	0	0	0	0	7	0	0	2	0	2	0	214	0	13	4	0	0
13	9	18	26	6	0	0	0	0	2	25	34	0	400	0	0	0	0
14	0	0	0	8	21	22	4	3	5	1	0	11	0	238	0	1	0
15	0	0	0	0	0	0	3	0	0	0	0	3	0	0	189	14	11
16	0	0	0	0	0	0	0	34	0	0	0	1	0	1	4	170	0
17	0	0	0	0	0	3	14	19	0	0	9	0	0	0	0	5	157

NUMBER CORRECT 5319 PERCENT CORRECT 76.6

2 = Assigned to treatment where result was (0) instead of correct one which would have been (++.)

or

Assigned to treatment where result was (-) instead of correct one which would have been (+.)

1 = Assigned to treatment where result was (+) instead of correct one which would have been (++.)

or

Assigned to treatment where result was (0) instead of correct one which would have been (+.)

or

Assigned to treatment where result was (-) instead of correct one which would have been (0.)

or

Assigned to treatment where result was (--) instead of correct one which would have been (-.)

0 = Assigned to treatment where result is the same.

The result is therefore a change degrading scale from 4 (maximum amount of degrading) to 0 (minimum amount of degrading.) By assuming that cases were equally distributed among tied "best treatments," we calculated the degrading effects of misclassification for the discriminant function approach. Of the 1,157 cases, 857.83 were cases in which misclassification results in no degrading of change. The overall effect is shown by type and overall in Table 29. This indicates that the overall mean degrading would be just under one half of a position, with a range from two full points to zero.

However, this mean could occur from a number of different combinations, some of them far more serious from a severity of misclassification point of view. For the 299.17 non-zero misclassifieds, the distribution is as indicated in Table 30. The overwhelming majority of degraded cases represent.

TABLE 29

MEAN CHANGE DEGRADING FOR MISCLASSIFIEDS BY TYPE

(Discriminant Function Method)

Type	Mean Degrading	N
1	2.00	2
2	.88	56
3	.19	309
4	.59	137
5	.00	48
6	.03	77
7	1.25	154
8	.00	18
9	.13	68
10	N/A	N/A
11	.19	105
12	1.36	28
13	.00	86
14	.02	58
15	1.00	11
16	N/A	N/A
17	N/A	N/A
Total	.40	1157

TABLE 30

INCIDENCE OF DEGRADING BY SEVERITY

(Discriminant Function Method)

Form of Misclassification		N	Percent
Misclassified as	Should Have Been		
--	++	0	0
--	+	17	6
-	++	3	1
-	+	8.83	3
0	++	57.67	19
0	+	148.33	50
-	0	1	0
+	++	61.34	21
--	-	2	1

a tendency toward "no change" misclassification, where positive change should have occurred. In summary form, 89 percent were instances of no positive change, whereas only 11 percent were instances of destructive change. Stated more succinctly, had a discriminant function approach been used, only one case in nine would have resulted in destructive consequences.

By Decision Tree Method

For the decision tree method, accuracy of classification is displayed in Table 31. Once more for the total of 6,944 cases, 4,972 (72 percent) were classified correctly. One thousand nine hundred seventy two cases (28 percent) were misclassified.

Of the 4,972, 4,804 cases were in other than Type 16, for which no treatment evidence was available in the original analysis. The same change-coding system (4 to 0) used in analyzing the discriminant function approach was applied to these cases. The mean change for these 4,804 correctly classified cases was, in this instance, 3.03 (once more "+.")

Of the 1,972 misclassified cases, 1,458 were in types (correct or misclassified) for which both correct classification change codes and misclassification change codes exist. Of the 1,458, 905.17 were cases in which misclassification results in no degrading of change. The overall effect is shown by type and overall in Table 32. This shows that the overall mean degrading would be just over one half of a position, with a range from slightly more than one and one half points to zero.

Again, this mean would occur from many particular combinations of differing severity. For the 552.83 non-zero misclassifieds, the distribution is presented in Table 33. Once more the majority of degraded cases

TABLE 31

THE COMPARISON OF CLASSIFICATION
BY THE
DISTANCE FUNCTION (ROWS) AND DECISION TREE (COLUMNS)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	301	16	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2	82	268	7	0	0	0	0	0	4	2	0	0	14	0	0	0	0
3	19	94	482	43	3	0	0	0	83	39	5	0	70	0	0	0	0
4	0	0	35	344	31	8	1	0	56	40	45	2	3	23	0	0	0
5	0	0	4	36	270	29	22	0	3	57	43	27	2	30	1	0	0
6	0	1	2	4	6	304	15	10	2	2	30	4	0	35	0	1	3
7	0	0	0	0	11	30	385	48	0	8	34	42	0	7	40	2	24
8	0	0	0	0	0	5	5	300	0	0	1	5	0	3	21	53	21
9	29	3	3	6	1	2	1	0	210	1	7	0	6	1	0	0	0
10	4	1	4	2	0	0	1	1	1	270	9	6	9	2	4	0	1
11	3	1	3	4	2	18	3	5	4	7	444	0	18	1	5	1	14
12	2	3	1	0	2	0	0	3	0	8	1	195	1	14	11	1	0
13	48	5	7	3	1	1	0	0	0	9	22	0	422	0	1	0	1
14	3	0	1	4	2	27	6	6	7	3	3	18	0	232	0	2	0
15	0	0	0	0	1	0	0	3	1	2	1	3	0	0	192	6	11
16	0	0	0	0	1	2	1	9	0	3	1	3	0	0	9	168	13
17	0	1	0	0	0	2	1	3	0	0	4	1	0	0	3	7	185

NUMBER CORRECT 4972 PERCENT CORRECT 71.6

TABLE 32

MEAN CHANGE DEGRADING FOR MISCLASSIFIEDS, BY TYPE

(Decision Tree Method)

Type	Mean Degrading	N
1	1.00	1
2	.77	109
3	.69	356
4	.41	244
5	.00	64
6	.39	114
7	.81	236
8	.00	40
9	1.57	60
10	.00	4
11	.68	71
12	.69	46
13	.00	39
14	.10	59
15	.73	15
16	N/A	N/A
17	N/A	N/A
Total	.59	1458

TABLE 33

INCIDENCE OF DEGRADING BY SEVERITY

(Decision Tree Method)

Form of Misclassification			
Misclassified as	Should Have Been	N	Percent
--	++	0	0
--	+	58	10
-	++	30.5	6
-	+	42.17	8
0	++	96.16	17
0	+	213.17	39
-	0	6	1
+	++	105.83	19
--	-	1	0

represent a tendency toward "no change," rather than toward outright deterioration.

However, the consequences are in this instance somewhat more severe. Seventy-five percent were instances of no positive change, while 25 percent were instances of destructive change. Stated otherwise, by a decision tree method, one case in four misclassifieds with degrading would have resulted in destructive change.

A Comparison to Change Agent Choice

An interesting question is the way in which either of these methods compare with the most common form of change treatment selection: change agent judgment or choice. To assess this, we turned to the data used in the original study which generated the hypothesis (Bowers & Hausser, 1977). Table 34 shows appropriate treatments, as determined by the data in that original analysis, proportions by type correctly classified and misclassified, plus, where feasible, a test of the difference between proportions.

Several observations seem worth making from these data. First, if assigned treatment is taken as indicative of classification, it is apparent that change agents were exceedingly poor judges, since they misclassified significantly more frequently than they classified correctly. Second, the differences by type are rather marked. Even ignoring types with small frequencies, the correct classification percentages range from 95 (for Type 14) down to three (for Type 3). Third, the misclassification discrepancies are most marked in the case of those types whose appropriate treatment was survey feedback alone.

This is demonstrated more pointedly in Table 35, where cases from those types calling for survey feedback alone and non-survey feedback alone are compared. Only one-fourth of the cases which should have called

TABLE 34

EFFECTS OF CHANGE CLASSIFICATION

Type	N	Correct Treatment*	Percent Correctly Classified	Percent Misclassified	Z	p
1	10	DH	70	30	-3.35	N/A
2	58	SF	28	72	-3.35	.001
3	98	SF	3	97	-9.31	.0001
4	53	SF	28	72	-3.20	.001
5	9	IPC	100	0		N/A
6	59	SF/IPC	34	66	-2.46	.01
7	41	SF	37	63	-1.66	NS
8	14	SF/IPC	100	0		N/A
9	26	SF	38	62	-1.22	NS
10	4	DH	100	0		N/A
11	27	SF	52	48	+0.02	NS
12	46	SF/TPC	41	59	-1.22	NS
13	11	SF/LT	100	0		N/A
14	56	IPC/TPC/LT	95	5	+6.73	.0001
15	17	IPC	35	65	-1.24	NS
16	--	(NA)	--	--		--
17	4	LT	100	0		--
Total	533		40	60	-4.03	.0001

*DH = Data Handback
SF = Survey Feedback
IPC = Interpersonal Process Consultation
TPC = Task Process Consultation
LT = Laboratory Training

TABLE 35

PERCENTS RECEIVING APPROPRIATE AND INAPPROPRIATE TREATMENT
FOR SURVEY FEEDBACK ONLY AND OTHER ONLY

Appropriate Treatment	Received Appropriate Treatment		Received Inappropriate Treatment	
	N	%	N	%
Survey Feedback only	73	24	230	76
Other than Survey Feedback Only	81	81	19	19
Z	-8.00			
p	0.0001			

only for survey feedback were assigned to that treatment, whereas more th three-fourths of the cases calling for some other treatment did, indeed, receive that other treatment.

Other data, not elaborated here, provide a bit more insight into this shift. Among the 76 percent of survey feedback only appropriate groups, which received some other treatment, only 15 percent received data handback Sixty-one percent therefore were assigned to some form of more process-oriented treatment (IPC, TPC, or LT), a shift which has been noted more broadly in anecdotal terms elsewhere (Bowers, 1976).

With this in mind, it seems useful to speculate on the dynamics of what might have occurred. First, it is important to note that, at the time in which these projects occurred, normative data were not routinely available, percentile scores were not provided, and the typology and treatment findings had not been generated. The change agents involved were therefore "flying blind," except for intuitive judgments that might have been based upon raw, tabulated survey data. Nevertheless, the classification process that occurred could, in light of what we now know, have made one or both of two kinds of errors:

- A misjudgment of level (e.g., viewing what was an I-60-65 profile as an I-45 profile)
- Ignoring, altering, or changing the weight given one or more components (e.g., viewing an I-60-65 profile as one with climate at the 60 level, but intra-group behaviors and processes at the 40 level.)

Both of these may well have occurred in the case of the 41 Type 7 groups, for example. This type, a straight-line or "I" profile at the 30-35 percentile level, called for survey feedback. Despite this, 20 groups (49 percent) received laboratory training, and six groups (15 percent)

received interpersonal process consultation. As an instance of the first type of error, it may have happened that the change agents reduced dissonance from what were their preferred styles by changing the level to either an I-40 or an I-25, both of which call for interpersonal process consultation. Alternatively, they may have subjectively chosen to focus on processes inside the groups and changed the level of organizational climate upward (to a Type 14), which calls for either of the process consultation treatments or laboratory training.

Whatever the dynamic, the data seem to demonstrate that the classification process implicit in treatment selection was far worse than would have been the case had either a multiple discriminant function or decision tree approach been used. It was on the whole even worse than chance. The unanswerable question, of course, is how well the change agents would have done, had they been provided with percentile scores, type designations, and treatment information geared to the typology.

Non-Accuracy Based Criteria

Information Required to Make a Decision

If a particular classification procedure requires considerably less information than another, yet performs (approximately) as well, it is more advantageous in that collecting and processing information is costly in terms of time and money. Different conceptual approaches to the reduction of data are discussed by Bowers, et al. (1977). The implementation of these procedures, where possible, are discussed herein.

One approach to reduce the amount of data used would be the deletion of some of the presently used 14 SOO indexes. One proposed technique involved an analysis of the discrimination capabilities of different indexes

within the Bayesian framework, i.e., those indexes with likelihood functions that do not vary greatly among profile types. The inability to implement the Bayesian approach, as discussed earlier, precluded such investigation. Also, the implementation of the decision tree algorithm took on a different form than originally proposed. The present algorithm treats all indexes equally. The prior concept was to sequence the indexes in order of importance. This would have permitted the comparison of classification of various (ordered) subsets of indexes with respect to accuracy. The design of the decision tree algorithm as used in this research, for reasons other than features discussed here, took an alternative form (unsequenced indexes.) Thus, no numerical investigations of the effect of deleting some indexes took place. However, this concern is still relevant, and ought to be considered in future classification investigations.

Another approach to data reduction is to reduce the number of indexes by combining them into "super-indexes," (rather than deletion.) This, in effect, is what was done to implement the discriminant function classification procedure. The inability to do the discriminant functions on the original set of 14 indexes, however, precludes examining the effect of this consolidation.

A final approach to data reduction is the elimination of some individual data from the data set. In many organizational settings, it is desired (usually by the organization) to survey only a sample of its employees. The question is what effect will this sampling have on the diagnostic process. If classification of work groups (or whatever else the primary unit of analysis might be) can be done with only a sample of employees, significant savings would occur.

A question that ought to be answered prior to such an investigation is, "What would be a satisfactory percentage agreement in classification between the total population and using a sample?" If, in fact, one could classify work groups using only a sample of employees with 98 percent agreement with what one would obtain using all employees, the sampling process would be indicated. Likewise, if the agreement is only 10 percent, one would not want to sample. Where is the dividing line? There is no factual information which indicates what one ought to choose for this acceptable level of agreement. The opinions of the present investigators were that an acceptable level of agreement would be 80 percent (with a little room for give, say down to 75 percent.) This was then used as the criteria of acceptability in sampling, i.e., that 80 percent or more of all work groups would have to be classified the same when using the sample data set as when using the entire data set.

A second feature of this investigation has to do with the sampling procedure used. With the Survey of Organizations, the primary unit of analysis is the work group. Thus, what is being classified is the work group. Yet, we are not sampling work groups, but individual employees. One natural approach would be to do a stratified sampling procedure, taking the same percentage of employees from within each work group. However, as the support for sampling is usually from the organization, their concerns ought to be considered here as well. Our experience is that the organization is usually interested in taking a sample of all employees without regard to work group affiliation. The effect of this is to have varying proportions of employees within each work group, with no assurance of a particular work group having any representation in the sample. While this may represent the worst possible situation with respect to diagnosis and

research, it has viability as being within real world constraints. For this reason, in an attempt to investigate the effect of sampling employees on classification, unstratified samples were selected.

Two different sample sizes were chosen for investigation; 67 percent and 33 percent. The procedure was to randomly select 67 percent (33%) of all employees, reform work group scores with the available data, and then to classify the work groups using the distance function. The resulting work groups which had only one individual's scores were omitted (as were those very large work groups.) The original number of classifiable work groups (with size between two and 40) was 5,651. For the 67 percent sample, there were 4,757 work groups, a reduction of 15.8 percent. In the 33 percent sample, there were 2,914 work groups, a reduction of 48.4 percent. These reductions are understandable in light of the fact that of the original 5,651 groups, 16.5 percent, 15.7 percent and 12.9 percent were of size two, three, and four, respectively.

The results of the sampling effects on classification are reported in Tables 36 and 37. The percentage of classification agreement is 65.1 percent and 42.7 percent for the 67 percent and 33 percent samples, respectively. These figures are well below the pre-established acceptable rate of 80 percent. The implication of this is that sampling does negatively impact organizational diagnosis at the work group level.

It is possible that had the sampling been done in a stratified manner (say, 67 percent of all work groups), the percent agreement would have been higher. It is unlikely, however, that it would be 15 percent higher. This does indicate further investigation. Other sampling plans might also be investigated. In particular, one might sample only from those work groups of size four or larger.

TABLE 36

CLASSIFICATION AGREEMENT USING A 67% SAMPLE

		Total Sample																
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	N= 4757	111	218	638	492	438	307	513	297	164	189	346	169	279	235	149	89	123
67% SAMPLE (1)	157	98	35	16	1	1	0	0	0	1	0	0	0	4	1	0	0	0
(2)	225	10	136	50	5	3	0	0	0	3	5	0	0	11	2	0	0	0
(3)	603	2	27	433	65	17	3	1	0	18	11	7	0	15	4	0	0	0
(4)	439	0	2	58	265	45	10	3	0	20	5	10	1	4	16	0	0	0
(5)	342	0	2	7	41	211	13	31	1	0	9	5	5	5	11	0	0	1
(6)	286	0	0	1	27	18	179	26	8	1	0	15	0	0	11	0	0	0
(7)	498	0	0	0	6	53	32	320	29	0	7	17	4	1	8	11	3	7
(8)	300	0	0	0	0	4	26	44	199	0	0	0	2	0	6	5	9	5
(9)	188	0	6	33	21	2	3	0	0	112	0	4	0	0	7	0	0	0
(10)	201	0	6	8	10	18	0	6	0	0	122	7	9	11	1	3	0	0
(11)	359	0	0	4	15	20	19	20	2	4	6	247	2	15	0	0	0	5
(12)	179	0	0	2	1	16	1	13	6	0	7	0	123	0	7	3	0	0
(13)	297	1	4	21	10	4	0	1	0	1	15	25	0	213	0	1	0	1
(14)	255	0	0	4	24	24	13	10	5	4	0	0	11	0	159	0	1	0
(15)	164	0	0	0	0	1	0	20	8	0	2	0	6	0	0	117	5	5
(16)	124	0	0	0	0	1	4	5	30	0	0	0	5	0	2	6	68	3
(17)	140	0	0	1	1	0	[illegible]	13	9	0	0	0	1	0	0	3	3	96

NUMBER CORRECT = 3098 PERCENT CORRECT = 65.125%

TABLE 37

CLASSIFICATION AGREEMENT USING A 33% SAMPLE

33% SAMPLE

		Total Sample (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
N=	2914	50	122	408	307	287	194	343	181	93	114	192	108	154	149	94	43	75
(1)	120	39	28	40	2	1	0	0	0	2	2	0	0	6	0	0	0	0
(2)	157	8	60	54	8	5	0	0	0	4	6	0	2	7	3	0	0	0
(3)	342	0	17	175	59	17	2	2	0	22	9	11	3	14	11	0	0	0
(4)	214	0	1	42	87	28	13	6	0	9	2	7	0	4	15	0	0	0
(5)	212	0	1	8	32	79	19	28	2	0	6	11	10	2	9	4	1	0
(6)	171	0	0	6	12	18	63	22	13	3	1	18	3	2	8	0	1	1
(7)	262	0	0	1	3	47	17	126	24	0	3	13	8	1	5	5	0	9
(8)	189	0	0	0	1	6	22	48	86	0	0	0	4	0	9	4	6	3
(9)	151	1	8	34	32	5	8	1	0	49	0	5	0	2	6	0	0	0
(10)	134	0	2	9	11	17	2	7	2	0	61	7	5	6	1	3	0	1
(11)	215	0	0	4	16	16	18	29	1	2	7	92	0	22	1	1	0	6
(12)	125	0	1	5	3	17	3	15	4	0	6	0	48	0	12	10	1	0
(13)	196	2	4	24	13	8	3	9	0	1	10	24	0	88	0	0	0	0
(14)	141	0	0	6	23	15	19	4	6	1	0	0	4	0	63	0	0	0
(15)	114	0	0	0	0	1	0	20	11	0	0	0	17	0	0	55	5	5
(16)	96	0	0	0	0	4	4	11	25	0	1	0	[illegible]	0	5	10	27	5
(17)	85	0	0	0	5	3	1	15	7	0	0	4	0	0	1	2	2	45

NUMBER CORRECT = 1243 PERCENT CORRECT = 42.656%

Another response to the results discussed above might be the suggestion to take a larger (than 67 percent) proportion of employees, say 80 percent. This would increase the agreement in classification. However, the benefits of using a sample rather than the entire population are greatly diminished. With that large of a proportion of the employees, the time/cost savings are minimized, the sampling costs are increased, and the questioning of the non-participation of only a few by the employees themselves is increased significantly. In originally choosing the 67 percent level, the authors consider it to be (at or near) the maximum level for sampling, given both organizational constraints and employee responses.

It was decided not to explore the discriminant function and decision tree classification algorithms on the 67 percent and 33 percent samples. As seen in the previous section, the accuracy of classification of these techniques, based on full data, is below 80 percent. The ability to reproduce the "correct" classification would be lower when using a fractional data set. A note of caution: a technique may be able to reproduce itself with greater accuracy on the sample data than the distance function procedure. What is being said here is that a technique will not reproduce the distance classification when using a partial data set than when using the total data set.

In summary, using samples of employees from an organization, rather than surveying all employees, appears to significantly alter the classification of work groups, and thus cannot be recommended for use in an organizational diagnosis project.

Amount of Data Required to Develop the Diagnostic Process

The three classification techniques require differential data bases for their development. The purpose of this section is to discuss available information regarding the necessary developmental data bases.

As earlier reported, the distance function technique, when applied to the total data set, generated clusters of work groups which represented the typology with a fairly high degree of compatability. However, there is no need for any data, other than what may be used to generate the typology, before implementing the distance function technique. That is, if one was to receive a typology from an outside source, no developmental work would be necessary.

However, in most instances, the typology will have to be developed. In the present case, it has been shown that a typology based on a developmental sample of N=600 was generalizable to the total data set (N=6,944).

The decision tree algorithm was based on intervals surrounding the means of indexes of the types. If one had intervals provided, no developmental data would be necessary. However, as before, this is unlikely. To obtain the intervals, one could easily take the means and standard deviations of the types resulting from the development of the typology. The stability of the intervals used in the decision tree algorithm would be of concern (in addition to the validity of the typology). To insure that the sample mean is close to the population mean, one would need samples of approximate size 400, 1,600, 6,300 for the standard error of $\overline{X}$ to be .02, .01, and .05, respectively, for each of the various indexes within types.

The discriminant function technique requires not only the typology, but an external "correct" classification on a developmental sample. The size of the developmental sample is complicated by the relative frequency of the various types. The smallest relative frequency times the sample size ought to be at least 50 or more (the greater the number of predictors, the larger this should be).

The results of the discriminant function analysis given earlier do have positive implications. The ability of the discriminant function to classify the test sample with the same accuracy as the developmental sample indicates the stability and transferability of the discriminant equations from a developmental sample to new data.

The conclusions of these remarks on the amount of information required to develop the classification procedures are presented in Figure 4.

Figure 4

Requirements for Development of
Classification Techniques

	Typology	Some Data Classified Into Types
Distance Function	x	
Decision Tree	x	x
Discriminant Function	x	x

Ease of Calculation

The ideal diagnostic technique is not only accurate, but relatively simple to perform. In an organizational diagnosis situation, the diagnostician may often be a change agent who would like to make treatment recommendations while on site and without having to resort to accessing and/or experience the time delay of electronic data processing (EDP) facilities.

Three ease-of-calculation factors are:

(1) Can be calculated on-site versus in a central location.

(2) Can be done by hand or with a hand calculator versus requiring EDP facilities.

(3) The amount of personnel time involved is relatively small.

While not identical, these three factors are clearly interrelated.

While analyzing the different classification techniques with respect to these issues is done subjectively, the difference between the techniques is relatively clear.

Both the distance function and the discriminant function techniques are clearly most easily done by computer. The distance function technique, however, given a table of index values for the types, could be implemented with a hand calculator (preferably one with accessible storage and minimal statistical routines) in 15 to 20 minutes (per unit of analysis, e.g., work group) by anyone with minimal training in the technique. The decision tree algorithm is clearly the most straightforward, least cumbersome technique. Anyone could use the technique to classify a new work group in five minutes or less -- without the aid of anything but a table of intervals and a pencil. The number of calculations for the decision tree algorithm is the smallest (14 x 17 yes, no questions), second lowest for

the discriminant function (although quite complex), and highest for distance function (although relatively simple arithmetically).

In summary, the decision tree methodology is the most portable, easily applied, and requires the minimal personnel time. The distance function can be done on-site, with a hand calculator, at a moderate level of complexity. The discriminant function, for reasonable application, would require programmable equipment with over 200 storage units.

DISCUSSION AND CONCLUSIONS

We began this research with a view that organizational diagnosis, as practiced in most development efforts, is largely unsystematic, if not capricious. We proposed to use a work group diagnostic typology as a basis for comparing several alternative, data based methods of performing the diagnostic task. As such, the questions which we wished to investigate revolved around the durability of the original typology and the effectiveness of the distance function approach upon which it was based, the accuracy with which discriminant function, Bayesian, and decision tree methods would classify groups, the ability to employ reduced information sets in diagnosis, and the comparative ease of classification of the various methods. The previous sections of the report have presented the results of our research; in the present section it remains to discuss their meaning and significance.

Durability of the Typology and Ability of the Techniques to Reproduce it

The typology does, indeed, appear to be quite durable. It holds up quite nicely when applied to the entire population of groups. While the Bayesian method met a dead-end early on, the remaining methods -- distance function, discriminant function, and decision tree -- all produce within-cluster variances which are not only not greater than, but are actually smaller than, within-cluster variances in the original developmental study's sample. This is significant because, had the original been a statistical or sampling artifact, we would have expected applying the distance function (or any of the other methods) to the entire population to produce in the fit-forcing within-cluster variances much larger values than those obtained

in the original study. The fact that, if anything, the reverse happens provides rather persuasive evidence that the typology really exists and represents more than an artifact, sampling or otherwise, of the methods used in the developmental study.

Beyond this, on the issue of size of within-cluster variances, no method is decidedly superior to the others. All produce variance values near or below those in the developmental study. However, as might perhaps be expected, the distance function produces in general cluster mean values closer to our criterion of the "true" values (those of the original study) than do either the discriminant function or decision tree methods.

Although we have rather arbitrarily chosen the cluster means of the original study as representing the time values, the problem is no doubt one of successive iterations, attempting each time to come closer to the mark. This provides us with a somewhat broader perspective on the problem.

That the typology does seem quite durable suggests rather strongly that there are true types, with true cluster values in 14-space, which exist. Viewed in somewhat factor analytic terms, there are true vectors, to which we are attempting in our research to rotate clusters classified by various methods. The variance (of groups in the population in 14-space) can thus be broken down into two components:

(a) Variance between the true vectors and mean cluster vectors.

(b) Variance within-clusters of classified cases from their cluster means.

If each method aligned its cluster vectors perfectly, they (the methods) would differ in accuracy only by within-cluster variance. As we have seen, the three methods do not differ appreciably on this. As a general

statement, therefore, we might say that the evidence for the typology is quite strong, that each method produces vectors which are equally "tight," but that the discriminant function and decision tree methods produce cluster mean values which are somewhat more distant from the original values than are those of the distance function. Since the latter may in some measure reflect the fact that the same method was used, great significance ought perhaps not be attached to it.

Problems in Implementing a Bayesian Approach

The Bayesian method met, within the limits of this study, a dead end because the profile indexes are not independent. Pursuing it despite this fact produced values for the odds-likelihood ratio, which exceeded the capacity of what is an up-to-date computer (AMDAHL 470V6). While a Bayesian approach cannot be permanently dismissed, its feasibility clearly requires that some answer be found to the problem of index interdependence. This interdependence is exacerbated when one has a fairly large set of predictors (say more than 7). It seems highly unlikely, however, that measures of organizational functioning will ever be independent, and most likely, will continue to be many in number.

Proportion of Correct Classification

The remaining alternatives to the distance function itself -- namely, discriminant function and decision tree -- seem on the surface to produce approximately equal proportions of correct classification (77 versus 72 percent.) However, this is true only if ties in the decision tree procedure that involve the correct type are assigned to that correct type. Omitting tied cases, classification accuracy for the decision tree drops to 57 percent.

The decision tree procedure, as implemented herein, produces many tied classifications. Here, as in its misclassifications, it seems unable to reliably distinguish straight-line profiles from divergent profiles at approximately the same level and tends to consign the former over much to the latter.

The tied classification problem for the decision tree method led us to apply, as an after thought, the severity calculation procedure (cf. p. 102) applied to misclassifications, to the ties themselves. In other words, perhaps the decision tree's tendency to ties is not, in its effects, much of a problem. Table 38 presents the results of this further analysis. They reveal that the problem is not appreciably different from that of misclassification. A large majority represent instances of an absence of positive change. However, again as with decision tree misclassifications, approximately one case in four or five would result in destructive change.

In part, of course, the ties and misclassification problem reflects the version of a decision tree selected for implementation in the present study. In theory, we might alternatively have chosen a sequenced form, in which being high on the first index in the sequence excludes some number of possibilities altogether. In fact, however, this possibility seems to be excluded by the state of the knowledge base. While there are sets of variables known on the basis of research to be causal in relation to other variables, the sequence is not fixed nor immutable. There is instead more nearly a gestalt flavor to the pattern of variables, with some "usually" or "generally" thought to precede others. Therefore, something like the present version of a decision tree seems the only feasible form, and, as we have seen, its accuracy leaves something to be desired.

TABLE 38

CHANGE DEGRADINGS FROM DECISION TREE TIES

Tie Leads To	Should Have Been	Amount of Degrading	N	Percent
--	++	4	0	0
--	+	3	42	2
-	++	3	12	1
-	+	2	69.5	4
0	++	2	115	6
0	+	1	221.16	12
-	0	1	3	0
+	++	1	130	7
--	-	1	5	0
No Degradings		0	1255.33	68

Zero-One Count

The three methods -- distance function, decision tree, and discriminant function -- do not differ appreciably on a zero-one count comparison. At roughly .67 standard deviation units, approximately half of the indexes per work group exceed the designated value of the assigned type by that much or more. Overall accuracy of classification notwithstanding, what this suggests is considerable variation, perhaps idiosyncratic, perhaps for particular indexes. However, the average number of indexes differing significantly from the designated value exceeds only slightly that which would be explained by random chance according to normal distribution theory.

Severity of Misclassification

The methods differ rather substantially in the consequences of their misclassification. While the consequences of correct classification under either the discriminant function or decision tree procedure would be a positive improvement of one quarter to one half a scale point, and the majority of misclassifications under either would result simply in an absence of positive change, the decision tree would much more frequently produce destructive change (outright deterioration) in its misclassifieds. Under the decision tree method, one misclassification in four would produce destructive change, whereas only one in nine would do so using a discriminant function. Either method would appear superior to change agent choice, however, since the latter produced results significantly worse than chance.

Use of Reduced Information Sets

Our results indicate that classification accuracy suffers rather dramatically when respondent sampling occurs. Even a two-thirds random sample, which would generally represent a higher proportion than is considered where sampling is proposed, produces a reduction in classification accuracy that is unacceptable.

However, at least a portion of the results suggest that reduced numbers of measures might produce equally satisfactory results. Our generation of "super-indexes" for use by the discriminant function, together with the fact that fewer than the number of items involved in each would presumably be required to obtain adequate internal consistency, suggests that it might be possible to generate a "short form" of the survey which would satisfy the diagnostic classification purpose. It should be noted, however, that that might not be compatible with feedback and development purposes.

Ease of Calculation

Only the decision tree method provides promise of easy calculation. Were its accuracy and severity of misclassification problems to be solved, therefore, it might present significant potential for being transportable and not bound to a large data bank and computer capability.

Conclusions from the Research

We conclude from this rather complex research project that the typology which undergrids our efforts at data based (computer assisted) diagnosis is indeed durable. Of the methods proposed for examination, two seem not to be likely candidates at any immediate time for on-line use in real world

settings. A Bayesian system, to be at all feasible, must first solve the problems of dealing with interrelated measures. This is no mean feat, since it seems highly likely that measures of organizational functioning are intrinsically inter-related, for a whole host of systemic reasons. Any effort to solve the problem by orthogonal factor analytic treatment of measures, for example, seems likely to encounter the difficulty that such efforts have in organizational settings in the past. What is orthogonal in one setting at one time is no longer so in other settings or at future times. Thus a Bayesian approach to organizational diagnosis seems improbable of realization without further research.

A decision tree approach, while eminently doable, as the present research has shown, seems to suffer from a mismatch between its requirements and the reality of the situation. The state of the knowledge base does not permit a decision tree of optimal form (i.e., a "sequenced" one). On the other hand, the fact that the non-sequenced form (like any other form) makes yes-no judgements means that it sacrifices information that both a distance and a discriminant function use. As a result, it makes errors of classification that are both too frequent and, more importantly, too costly in their likely impact upon change outcomes.

Even the decision tree method, however, and certainly each of the others examined, would yield results superior to change agent choice, which did significantly worse than chance.

Respondent sampling seems not to be a feasible route to making the survey-diagnostic task more manageable, since even relatively large samples produce unacceptable errors of accuracy. Collapsing of measures and reduction of the number of items does seem a potentially feasible route, however, provided that the task of the measures is limited to diagnostic classifications.

Perhaps one of the most significant findings is that neither of the comparatively accurate methods is easy of calculation. Only the decision tree method has this potential quality and its problems of accuracy clearly outweigh that ease. It seems likely, therefore, that any acceptable method of organizational diagnosis will for the foreseeable future require the amassing of a large data bank and the availability of a significant computer accessibility.

Against this backgrop, our conclusion is that research should be pushed on the following issues of organizational diagnosis, answers to all of which could greatly broaden the scope of what is possible and implementable.

- Additional exploration of the typology, including any possible modifications of the definitional values, its generalizability and durability.
- Enlarge the knowledge base with respect to the effects of different treatments on different types within the typology.
- Investigation of the effect of level on work group type. One question which illustrates this issue is whether a Type three, level one is the same as a Type three, level four. (Level refers to level of the work group in the organization.)
- How the accuracy of classification is affected by the level of the work groups being classified.
- The effect on the typology of converting work group scores to level percentiles prior to classification.

- Additional inquiry into the decision tree type of algorithm, with possible usage of non-uniform intervals, both across types and across indexes.
- Investigation into other forms of data reduction with respect to sampling. Possibilities might include stratified sampling or sampling only from work groups with size above a certain level.

SUMMARY

The process of organizational development and change contains two component subprocesses, diagnosis and therapeutic intervention. Although equally crucial to the success of any developmental effort, diagnosis takes prior importance because it occurs earlier in the developmental stream. Thus, an important part of the consultant's role is often presumed to consist of translating a wide variety of symptoms into a coherent pattern that permits planning and carrying out appropriate remedial action.

If this is the desirable state of affairs, it is scarcely what in fact obtains in most organizational development projects. Reviews and exchanges in the literature, plus the meager formal evidence, suggest that the field is characterized by ad hoc problem-solving and by efforts to simply justify whatever it is that the consultant knows how to do.

If progress is to be made in organizational diagnostic methodology, procedures must be explored and tested which offer promise of systematizing the task. In other fields with a similar diagnostic task, such as medicine and psychiatry, advances have been registered by computerizing portions of the process. In addition to speeding up the procedure, this offers promise of handling as well the general problem of the clinical judge (consultant): statistical prediction, here as elsewhere, is likely to prove far more accurate than clinical, or clinically mediated, prediction.

An attempt to develop and test more rigorous diagnostic procedures in organizational development should be based upon a model containing principles of organizational functioning. The Likert meta-theoretical statement

formed the model underlying the research herein reported, as well as the source of content of the Survey of Organizations' data bank, containing approximately 7,000 work groups, used to test the propositions which were investigated.

An earlier study had used a hierarchical grouping technique which employs the distance function as a measure of work group Survey of Organizations' profile similarily to generate a 17-type typology. The type profile means from this earlier study were taken as the patterns of symptoms defining the 17 alternative functional states, and the distance function was used with all 7,000 groups to define the "true" type of each group. Against this criterion, a number of questions were addressed:

- Is there evidence that the original typology holds up when applied to the entire population of groups; that is, does it appear to reflect reality?
- If the typology holds up, how effective is a distance function in classifying by type?
- Using the distance function classification as the criterion, how accurate are discriminant function, decision tree, and Bayesian methods in classifying groups?
- To what extent are these methods able to classify groups correctly using reduced information sets?
- Do the methods differ in ease of calculation?

The results indicated that the typology did indeed hold up, with within-cluster variances for all methods used to reproduce it which no larger than, and ordinarily smaller than, those of the original study. As might

be expected, the distance function produced profile mean scores which were closer to those of the original study than were those of the other methods examined.

The Bayesian method met with an early dead-end, since the survey indexes were both intrinsically and in fact interrelated. An attempt to calculate odds-likelihood ratios despite this fact produced values which exceeded the capacity of what is a very large computer. While this method ought not be permanently dismissed, further progress is obviously contingent upon some answer being found to the problem of index interdependence.

On the surface, the decision tree and discriminant function methods appeared to be approximately equal in their ability to classify groups correctly. However, the decision tree method, at least as implemented within this study, seemed prone to an inability to differentiate among types with some characteristics in common. Specifically, it generated an excessive number of ties. For both these ties and its outright misclassifications, there was a degrading of projected change which was quite severe: one out of four misclassified groups would experience outright destructive consequences, whereas only one in nine would do so using a discriminant function method. Any of the methods was superior to consultant choice, however, where results were significantly worse than chance.

Reduced information sets in the form of fewer measures seem entirely plausible from the study's results. On the other hand, sampling of respondents produced unacceptable decrements in classification accuracy even at high percentages.

Both the distance and discriminant function methods require large developmental data banks and a significant computer capacity for their

implementation. Only the decision tree method could stand alone without these supports. However, its problems of accuracy and of the consequences of its errors make that transportability a dubious advantage at present.

References

Birnbaum, A. & Maxwell, A.E. Classification procedures based on Bayes' formula. Applied Statistics, 1960, 9, 152-169.

Bowers, D.G. OD techniques and their results in 23 organizations: The Michigan ICL Study. The Journal of Applied Behavioral Science, 1973, 9 (1).

Bowers, D.G. Organizational Development: Promises, performances, possibilities. Organizational Dynamics, 1976a, 4 (4),51-62.

Bowers, D.G. Systems of organization. Ann Arbor, Michigan: University of Michigan Press, 1976b.

Bowers, D.G. Organizational diagnosis: A review and a proposed method. Technical Report to Office of Naval Research, 1974.

Bowers, D.G. & Hausser, D. Group types and intervention effects in organizational development. Technical Report to Office of Naval Research, 1975.

Bowers, D.G. & Franklin, J.L. Survey-guided development: Data based organizational change. LaJolla, Calif.: University Associates, 1977.

Bowers, D.G. & Hausser, D.L. Work group types and intervention effects in organizational development. Administrative Science Quarterly, 1977, 23, 76-93.

Bowers, D.G. & Spencer, G.J. Structure and process in a social systems framework. Organization and Administrative Sciences, 1977, 8 (1), 13-21.

Bowers, D.G., Davenport, A.S. & Wheeler, G.E. Comparative issues and methods in organizational diagnosis. Technical Report to the Office of Naval Research, November, 1977.

Burke, W. Organization development. Professional Psychology, 1973, 4, 194-200.

Cronbach, L.J. Essentials of psychological testing. New York: Harper & Row, Publishers, 1960.

Crooks, J., Murray, I.P.C., & Wayne, E.J. Statistical methods applied to the clinical diagnosis of Thyrotoxocosis. Quarterly Journal of Medicine, 1959, 28, 211-234.

Fleiss, J.L., Spitzer, R.L., Cohen, J., & Endicott, J. Three computer diagnosis methods compared. Archives of General Psychiatry, 1972, 27, 643-649.

Fox, D.J. & Guire, K.E. Documentation for MIDAS. University of Michigan, Statistical Research Laboratory, 1976, 3rd ed.

Franklin, J.L. Characteristics of successful and unsuccessful organization development. Journal of Applied Behavioral Science, 1976, 11, (4), 471-492.

Gustafson, D.H., Edwards, W., Phillips, I.D., & Slack, W.V. Subjective probabilities in medical diagnosis. I.E.E.E. Transactions on Man-Machine Systems, September, 1969, 10, (3), 61-65.

Holland, J.H. Adaptation in natural and artificial systems. Ann Arbor, Michigan: The University of Michigan Press, 1975.

Kahn, R.L. Organizational development: Some problems and proposals. Journal of Applied Behavioral Science, 1974, 10, 485-502.

Lawrence, P.R. & Lorsch, J.W. Developing organizations: diagnosis and action. Reading, Mass.: Addison-Wesley Publishing Co., 1969.

Ledley, R. & Lusted, L. Reasoning foundations of medical diagnosis. Science, 1959, 130, July 14, 9-21.

Levinson, H. The clinical psycholgist as organizational diagnostician. Professional Psychologist, 1972, 3, 34-40.

Levinson, H. Response to Sashkin and Burke. Professional Psychologist, 1973, 4, 200-204.

Likert, R.L. New patterns of management. New York: McGraw-Hill, 1961.

Likert, R.L. The human organization. New York: McGraw-Hill, 1967.

Likert, R.L. & Likert, J.G. New ways of managing conflict. New York: McGraw-Hill, 1976.

Likert, R.L. Past and future perspectives on System 4. Academy of Management remarks, 1977.

Meehl, P.E. Clinical versus statistical prediction. Minneapolis, Minn.: University of Minnesota Press, 1954.

Melrose, J.P., Stroebel, C.T., & Glueck, B.C. Diagnosis of psychopathology using stepwise multiple discriminant analysis. Comparative Psychiatry, 1970, 11, 43-50.

Michaelsen, L. A methodology for the studies of the impact of organizational values, preferences, and practices on the All-Volunteer Navy. Technical Report to Office of Naval Research, 1973.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1967.

Overall, J.E. & Gorham, D.R. A pattern probability model for the classification of psychiatric patients. Behavioral Science, 1963, 8, 108-116.

Overall, J.E. & Hollister, J.E. Computer procedures for psychiatric classification. Journal of the American Medical Association, 1964, 187, 583-588.

Overall, J.E. & Klett, C.J. Applied multivariate analysis. New York: McGraw-Hill, 1972.

Pecorella, P.A. & Bowers, D.G. Future performance trend indicators: A current value approach to human resources accounting. Technical Report to Office of Naval Research, 1976a.

Pecorella, P.A. & Bowers, D.G. Future performance trend indicators: A current value approach to human resources accounting. Technical Report to Office of Naval Research, 1976b.

Rao, C.R. & Slater, P. Multivariate analysis applied to differences between neurotic groups. British Journal of Psychology 2 (statistical section), 1949, 17-29.

Sashkin, M. Organization development practices. Professional Psychologist, 1973, 4, 187-194.

Sletten, I.W., Altman, H., & Ulett, G.A. Routine diagnosis by computer. American Journal of Psychiatry, 1971, 127, 1147-1152.

Smith, W.G. A model for psychiatric diagnosis. Archives of General Psychiatry, 1966, 14, 521-529.

Spencer, G.J. A methodology for the studies of the impact of organizational values, preferences, and practices on the United States Army. Technical Report to the U.S. Army Research Institute for the Behavioral and Social Sciences, 1975.

Spitzer, R.L. & Endicott, J. DIAGNO: A computer program for psychiatric diagnosis utilizing the differential diagnostic procedure. Archives of General Psychiatry, 1968, 18, 746-756.

Spitzer, R.L., Endicott, J., Cohen, J., & Fleiss, J.L. Constraints on the validity of computer diagnosis. Archives of General Psychiatry, 1974, 31, 197-203.

Taylor, J.C. & Bowers, D.G. Survey of Organizations: A machine-scored standardized questionnaire instrument. Ann Arbor, Michigan: Institute for Social Research, 1972.

Tichy, N. Agents of planned social change: congruence of values, cognitions, and actions. Administrative Science Quarterly, 1974, 19, 164-182.

Veldman, D.J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart & Winston, 1967.

Wing, J.K. A standard form of psychiatric present state examination and a method for standardizing the classification of symptoms, in Hare, E.H. & J.K. Wing (eds.), Psychiatric epidemiology: An international symposium. London: Oxford University Press, 1970, 93-108.

DISTRIBUTION LIST

Office of Naval Research
(Code 452)
800 N. Quincy St.
Arlington, VA 22217

Director
U.S. Naval Research Laboratory
Washington, DC 20390
ATTN: Technical Information Division

Defense Documentation Center
Building 5
Cameron Station
Alexandria, VA 22314

Library, Code 2029
U.S. Naval Research Laboratory
Washington, DC 20390

Science & Technology Division
Library of Congress
Washington, DC 20540

Commanding Officer
ONR Branch Office
495 Summer St.
Boston, MA 02210

Psychologist
ONR Branch Office
495 Summer St.
Boston, MA 02210

Commanding Officer
ONR Branch Office
536 S. Clark St.
Chicago, IL 60605

Commanding Officer
ONR Branch Office
1030 E. Green St.
Pasadena, CA 91106

Psychologist
ONR Branch Office
1030 E. Green St.
Pasadena, CA 91106

Research Psychologist
ONR Branch Office
536 Clark St.
Chicago, IL 60605

Director
Program Management
ARPA, Room 813
1400 Wilson Blvd.
Arlington, VA 22209

Director
OSD-ARPA R&D Field Unit
APO San Francisco 96243

Director
Cybernetics Technology Office
ARPA, Room 625
1400 Wilson Blvd.
Arlington, VA 22209

Dr. Earl A. Alluisi
Old Dominion University Res. Foundation
Norfolk, VA 23508

Dr. Judith Daly
Decisions & Designs, Inc.
Suite 100
8400 Westpark Dr.
McLean, VA 22101

Dr. James A. Bayton
Dept. of Psychology
Howard University
Washington, DC 20001

Dr. H. Russell Bernard
Dept. of Sociology & Anthropology
West Virginia University
Morgantown, West Virginia 26506

Dr. Arthur Blaiwes
Naval Training Equipment Center
Orlando, Florida 32813

Dr. Milton R. Blood
School of Business
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Davis B. Bobrow
University of Maryland
Dept. of Government & Politics
College Park, MD 20742

Dr. John J. Collins
Vice President
Essex Corp.
201 N. Fairfax St.
Alexandria, VA 22314

Dr. Faris Kirkland
University City Science Center
Center for Social Development
3624 Science Center
Philadelphia, PA 19104

Dr. C. Brooklyn Derr
Associate Professor, Code 55
Naval Post Graduate School
Monterey, CA 93940

Dr. George T. Duncan
Carnegie-Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213

Dr. William E. Gaymon
Richard A. Gibboney Associates, Inc.
10605 Concord St.
Kensington, MD 20795

Dr. Paul S. Goodman
Graduate School of Industrial Admin.
Carnegie-Mellon University
Pittsburgh, PA 15213

Dr. J. Richard Hackman
Administrative Sciences
Yale University
56 Hillhouse Ave.
New Haven, Conn. 06520

Dr. Leo A. Hazlewood
CACI, Inc.
1815 Fort Myer Dr.
Arlington, VA 22209

Dr. Edwin Hollander
Dept. of Psychology
State University of New York at Buffalo
4230 Ridge Lea Rd.
Buffalo, NY 14226

Mr. Daniel F. Huck
General Research Corp.
Westgate Research Park
McLean, VA 22101

Dr. Charles L. Hulin
Dept. of Psychology
University of Illinois
Champaign, Ill. 61820

Dr. Rudi Klauss
Syracuse University
Public Administration Dept.
Maxwell School
Syracuse, NY 13210

Dr. Edward E. Lawler
Battelle Human Affairs Research Centers
4000 N.E. 41st St.
P O Box 5395
Seattle, Washington 98105

Dr. Arie Y. Lewin
Duke University
Duke Station
Durham, N.C. 27706

Dr. Morgan W. McCall, Jr.
Center for Creative Leadership
5000 Laurinda Dr.
P. O. Box P-1
Greensboro, N.C. 27402

Dr. Terence R. Mitchell
School of Business Administration
University of Washington
Seattle, WA 98195

Dr. William H. Mobley
College of Business Administration
University of S. Carolina
Columbia, S.C. 29208

Dr. James P. Murphy
National Analysts
A Division of Booz-Allen & Hamilton, Inc.
400 Market St.
Philadelphia, PA 19106

Dr. Peter G. Nordlie
Human Sciences Research, Inc.
7710 Old Springhouse Rd.
McLean, VA 22101

Dr. Herbert R. Northrup
Industrial Research Unit
University of Pennsylvania
Philadelphia, PA 19174

Dr. A.F.K. Organski
3068 Institute for Social Research
University of Michigan
Ann Arbor, MI 48104

Dr. Manual Ramirez
Systems and Evaluations
232 Swanton Blvd.
Santa Cruz, CA 95060

Dr. Irwin Sarason
Dept. of Psychology
University of Washington
Seattle, WA 98195

Dr. S.B. Sells
Texas Christian University
Fort Worth, Texas 76129

Dr. H. Wallace Sinaiko
Program Director
Manpower Research & Advisory Services
Smithsonian Institution
801 N. Pitt St. - Suite 120
Alexandria, VA 22314

Mrs. Alice I. Snyder
1749 Navaja Lane
El Cajon, CA 92020

Dr. Richard Steers
Graduate School of Management & Business
University of Oregon
Eugene, Oregon 97403

Dr. Victor H. Vroom
School of Organizational Management
Yale University
56 Hillhouse Ave.
New Haven, Conn. 06520

Dr. Abraham R. Wagner
Analytical Assessments Corp.
P.O. Box 9695
Marine del Rey, CA 90291

Dr. J. Wilkenfeld
Dept. of Government & Politics
College Park, MD 20742

AFOSR/NL
Bldg. 410
Bolling AFB
Washington, DC 20332

Military Assit. for Human Resources
OAD (E&LS) ODDR&E
Pentagon 3D129
Washington, DC 20301

Air University Library/LSE 76-443
Maxwell AFB, AL 36112

Office of the Deputy Chief of Staff
for Personnel, Research Office
ATTN: DAPE-PBR
Washington, DC 20310

Chief, Plans & Operations Office
USA Research Institute for the
Behavioral & Social Sciences
Room 278
1300 Wilson Blvd.
Arlington, VA 22209

Army Research Institute (2 cys)
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. A.L. Slafkosky
Code RD-1
HQ US Marine Corps
Washington, DC 20380

Commandant of the Marine Corps
(Code MPI-20)
Washington, DC 20380

Chief, Psychological Research Branch
US Coast Guard (G-P-1/62)
400 7th St. SW
Washington, DC 20590

Chief of Naval Personnel
Asst. for Research Liaison (Pers-Or)
Washington, DC 20370

Bureau of Naval Personnel (Pers 6)
Asst. Chief of Naval Personnel for Human Resource Management
Washington, DC 20370

Bureau of Naval Personnel (Pers-6a3)
Human Resource Management Financial Office
Washington, DC 20370

Naval Postgraduate School
Monterey, CA 93940
ATTN: Lirary (Code 2124)

Professor John Senger
Operations Research & Administration Sciences
Naval Postgraduate School
Monterey, CA 93940

Scientific Director
Naval Health Research Center
San Diego, CA 92152

Navy Personnel R&D Center
San Diego, CA 92152

Commanding Officer
Naval Training Equipment Center
Technical Library
Orlando, FL 32813

Dr. C. Brooklyn Derr
Associate Professor, Code 55
Naval Postgraduate School
Monterey, CA 93940

Human Resource Management Center
Box 23
FPO, New York 09510

Human Resource Management Center, Norfolk
5621-23 Tidewater Dr.
Norfolk, VA 23511

Human Resource Management Center
Bldg. 304
Naval Training Center
San Diego, CA 92133

Office of Naval Research (Code 200)
Arlington, VA 22217

ACOS Research & Program Development
Chief of Naval Education & Training (N-5)
Naval Air Station
Pensacola, FL 32508

Human Resource Management Center
Pearl Harbor
FPO San Francisco, CA 96601

Human Resource Management School
Naval Air Station Memphis (96)
Millington, TN 38054

Capt. Bruce Stone, U.S.N.
Director, Programs Development Division
(Code N-35)
Chief of Naval Education & Training
Naval Air Station
Pensacola, FL 32508

David Timberlake
Bureau of Naval Personnel
Pers 65
Washington, DC 20370

Director, Human Resource Training Dept.
Naval Amphibious School
Little Creek
Naval Amphibious Base
Norfolk, VA 23521

Human Resource Management Center,
Washington
Washington, DC 20370

Ms. Elsa A. Porter
Clearinghouse on Productivity &
Organizational Effectiveness
U.S. Civil Service Commission
Washington, DC 20415

HumPRO (Attn: Library)
300 N. Washington St.
Alexandria, VA 22314

Canadian Defense Liaison Staff,
Washington
2450 Massachusetts Ave., NW
Washington, DC 20008
ATTN: Chief, Defense Research

Mr. Luigi Petrullo
2431 N. Edgewood St.
Arlington, VA 22207

Dr. Meredith Crawford
5605 Montgomery St.
Chevy Chase, MD 20015

CDR Donald F. Parker, USN
Management Dept.
U.S. Naval War College
Newport, RI 02840

Dr. Diane C. Ryan
McBer and Company
137 Newbury St.
Boston, MA 02116

Dr. William S. Maynard
Dept. of Leadership & Law
U.S. Naval Academy
Annapolis, MD 21402